UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M Working Papers
Statistics and Econometrics
16-11
ISSN 2387-0303
November 2016

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

A Bootstrap Approach for Generalized Autocontour Testing

Joao Henrique G. Mazzeu[a], Gloria González-Rivera[b], Esther Ruiz[a,c,*], Helena Veiga[a,c,d]

Abstract

We propose an extension of the Generalized Autocontour (G-ACR) tests (González-Rivera and Sun, 2015) for one-step-ahead dynamic specifications of conditional densities *in-sample* and of forecast densities *out-of-sample*. The new tests are based on probability integral transforms (PITs) computed from bootstrap conditional densities that incorporate the parameter uncertainty without assuming any particular forecast error distribution. Consequently, the parametric specification of the conditional moments can be tested without relying on any particular error distribution. We show that the asymptotic distributions of the bootstrapped G-ACR (BG-ACR) tests are well approximated using standard asymptotic distributions. Furthermore, the proposed tests are easy to implement and are accompanied by graphical tools which provide suggestions about the potential misspecification. The results are illustrated by testing the dynamic specification of the Heterogenous autoregressive (HAR) model when fitted to the popular U.S. volatility index VIX.

Keywords: Distribution Uncertainty; Model Evaluation; Parameter Uncertainty; PIT

[a] Department of Statistics, Universidad Carlos III de Madrid.
[b] Department of Economics, University of California, Riverside.
[c] Instituto Flores de Lemus, Universidad Carlos III de Madrid.
[d] Unide, Instituto Universitário de Lisboa.

Preliminary and incomplete. Do not cite.

Financial support from the Spanish Ministry of Education and Science, research project ECO2015-70331-C2-2-R (MINECO/FEDER) is acknowledged by the four authors.

* C/ Madrid, 126, 28903, Getafe, Madrid (Spain), Tel: +34 916249851, Fax: +34 916249849, Email: ortega@est-econ.uc3m.es. Corresponding Author.

1. Introduction

Density forecasting is rapidly becoming a very active and important area of research in the analysis of economic and financial time series. The need to consider the full predictive density has long been recognized in the related literature; see Tay and Wallis (2000) for a survey. There are several reasons for this growing interest in density forecasting. First, complete probability distributions over outcomes provide helpful information for making economic decisions; see Granger and Pesaran (2000a,b). Second, density forecasts provide a characterization of forecast uncertainty which can be useful to central banks; see Britton et al. (1998) for the fan charts of the Bank of England and Alessi et al. (2014) for measures of economic uncertainty during the Global Financial Crisis of the Federal Reserve Bank of New York. Soyer and Hogarth (2012) also propose incorporating measures of uncertainty to avoid the illusion of predictability. Third, in the presence of non-normal forecast errors, even single forecast intervals may not provide an adequate summary of the expected future; see, for example, Lam and Veall (2002). Fourth, density forecasts are also important in the presence of realistic economic loss functions which cannot be reduced to the comparison of Mean Squared Forecast Errors of point forecasts; see Diebold and Mariano (2002) and Patton and Timmermann (2007). Furthermore, in some applications, often the object of interest is a particular quantile of the forecast distribution as, for example, when forecasting the Value-at-Risk (VaR) of a given stock or portfolio; see Nieto and Ruiz (2016) for a recent survey on VaR forecasting. As a consequence, in an increasing number of empirical applications, forecast densities are obtained for macroeconomic and financial variables; see, for example, Fair (1980), for one of the first applications of computing probability forecasts using a macroeconomic model. Garratt et al. (2003), Giordani and Villani (2010), Jore et al. (2010), Clark (2011), Baumeister and Kilian (2012), Clark and Ravazzolo (2015) and Ravazzolo and Rothman (2016) are some more recent macroeconomic applications. The number of applications in the context of financial variables is very broad covering the construction of densities for both returns and volatilities; see, Andersen et al. (2003), Clements et al. (2008), Corradi et al. (2009), Maheu and McCurdy (2011) and Hallam and Olmo (2013, 2014) just to mention a few applications. Note that, in some of these applications, the forecasting densities are multivariate.

A problem often faced by forecasters is testing the correct specification of a conditional forecast density. Appropriate tests should take into account that the forecast conditional distribution is often unknown and the specification of conditional moments is also unknown and has estimated parameters. Furthermore, a useful test will indicate the source of rejection of a given forecasting model, that is, whether it is rejected because of the specification of the shape of the distribution or because of the specification of the conditional moments.

Many tests available in the literature are based on testing a joint hypothesis of uniformity and independence (i.i.d. U(0,1)) of the probability integral transforms (PITs), which are applicable regardless of the particular user's loss function. Among these tests, the most popular is due to Diebold et al. (1998); see also Berkowitz (2001) and Chen and Fan (2004) for extensions. Intuitively, the i.i.d. assumption of the PITs is related with the correct specification of the conditional moments, while the U(0,1) property characterizes the correct specification of the distribution. The PITs contain rich information on model misspecification which can be revealed by using their histogram and autocorrelogram as suggested by Diebold et al. (1998). However, none of these visual devices take into account the uncertainty associated with parameter estimation. Furthermore, it is nontrivial to develop a formal test for the joint hypothesis of independence and uniformity of the PITs. The well-known Kolmogorov-Smirnov test, checks uniformity under the independence assumption rather than testing both properties jointly. Consequently, it would easily miss the non-independent alternatives when PITs have a marginal uniform distribution. Moreover, the Kolmogorov-Smirnov test does not take into account the impact of parameter estimation uncertainty on the asymptotic distribution of the statistic. To solve this problem, Bai (2003) proposes a Kolmogorov-Smirnov-type test based on a martingale transformation of the PITs whose asymptotic distribution is free from the impact of parameter estimation. However, the test proposed by Bai (2003) only checks uniformity and, consequently, it has no asymptotic unit power if the transformed PITs are uniform but not independent; see Corradi and Swanson (2006). Alternatively, Hong and Li (2005) propose a nonparametric-kernel-based test with power against violations of both independence and density functional form. Nevertheless, it depends on the choice of a bandwidth and, consequently, it is problematic how to choose it in an empirical context.

Instead of testing for independence and uniformity of PITs, González-Rivera et al. (2011) and González-Rivera and Yoldas (2012) propose autocontour (ACR) tests to evaluate the adequacy of conditional forecast densities. Relying on autocontours allows to obtain a graphical tool that can be very helpful for guiding the modelling. Moreover, it permits to focus on different areas of the conditional density in order to assess those regions of interest. The ACR test, which can be applied to both original series and model residuals, has several advantages: i) it has standard convergence rates and standard limiting distributions that deliver superior power; ii) it is computationally easy to implement as it is based on counting processes; iii) it does not require either a transformation of the original data or an assessment of the Kolmogorov goodness of fit; and iv) it explicitly accounts for parameter uncertainty. Yet, it assumes a parametric time-invariant function of the forecast density and it is complicated to implement to multivariate forecast densities. To overcome these problems, González-Rivera and Sun (2015) propose the generalized autocontour (G-ACR) test, that is based on PITs instead of original observations or residuals. In this way, the G-ACR test inherits the advantages of using PITs and of using autocontours. However it is still based on assuming a particular specification of the conditional density in order to compute the PITs. Therefore, when a given forecasting model is rejected, it is difficult to disentangle whether the rejection can be attributed to the assumed functional form of the error distribution or to the specification of the conditional moments. González-Rivera and Sun (2015) point out that the G-ACR tests are more powerful for detecting departures from the distributional assumption than for detecting misspecified dynamics. Furthermore, there are applications in which the density does not have a known closed-form solution, as for example, multi-step predictive densities in non-linear or non-Gaussian models.

In this paper, we propose an extension of the G-ACR tests for dynamic specification of a density model (in-sample tests) and for evaluation of forecast densities (out-of-sample tests). Our contribution lies on computing the PITs from a bootstrapped conditional density so that no assumption on the functional form of the forecast error density is needed[2]. The only restrictions required on the error density are those needed to guarantee that the estimator of the parameters of the conditional moments is consistent and asymptotically Normal distributed. The bootstrap procedure allows for the incorporation of parameter uncertainty and can be extended to multivariate systems and multi-step forecasts. We show that the asymptotic distributions of the bootstrapped G-ACR (BG-ACR) tests are well approximated using standard asymptotic distributions. The proposed approach is very easy to implement and particularly useful to evaluate forecast densities when the error distribution is unknown. Furthermore, using graphical devices, the procedure

[2]Bootstrapping was also proposed by Tsay (1992) for model checking because of its flexibility. The essence of the procedures proposed by Tsay (1992) is to obtain the empirical distribution of a specified functional via parametric bootstrap, which then serves to compare the corresponding functional quantity. The spirit of the procedure proposed in this paper is very similar. However, differently from Tsay (1992), we do not assume known parameters or a known distributional assumption of the errors.

allows the identification of the source of misspecification, namely, whether, it is the error distribution, or linear or non-linear dynamics.

The rest of the paper is organized as follows. In section 2, we briefly describe the G-ACR test. Section 3 contains the main contribution of this paper with the description of the new proposed BG-ACR tests. Their asymptotic properties and finite sample performance are also analyzed when implemented in-sample. Section 4 is devoted to analyzing their out-of-sample behavior. An empirical application to illustrate the advantages of the BG-ACR tests, when implemented to test for the adequacy of the Heterogeneous Autoregressive (HAR) model to obtain forecast densities of the VIX volatility index, is carried out in section 5. Finally, section 6 concludes.

2. The G-ACR test

In this section, we briefly describe the G-ACR test proposed by González-Rivera and Sun (2015).

Let $\{y_t\}_{t=1}^T$ denote the random process of interest with conditional density function $f_t(y_t|Y_{t-1})$, where $Y_{t-1} = (y_1, ..., y_{t-1})$ is the information set available up to time t-1. Observe that the random process y_t might enjoy of very general statistical properties, e.g. heterogeneity, dependence, etc. A conditional model is constructed by specifying a conditional mean, conditional variance or other conditional moments of interest, and making distributional assumptions on the functional form of $f_t(y_t|Y_{t-1})$. Based on the conditional model, the researcher might construct a density forecast denoted by $g_t(y_t|Y_{t-1})$ and obtain a sequence of PITs of $\{y_t\}_{t=1}^T$ w.r.t $g_t(y_t|Y_{t-1})$ as follows

$$u_t = \int_{-\infty}^{y_t} g_t(v_t|Y_{t-1})\, dv_t. \tag{1}$$

If $g_t(y_t|Y_{t-1})$ coincides with the true conditional density, $f_t(y_t|Y_{t-1})$, then the sequence of PITs, $\{u_t\}_{t=1}^T$, must be i.i.d. $U(0,1)$; see Rosenblatt (1952) and Diebold et al. (1998). Therefore, the null hypothesis $H_0 : g_t(y_t|Y_{t-1}) = f_t(y_t|Y_{t-1})$ is equivalent to the null hypothesis

$$H_0' : \{u_t\}_{t=1}^T \quad \text{is i.i.d.} \quad U(0,1). \tag{2}$$

Note that, if the forecast density coincides with the true DGP, then it is preferred by all forecasters regardless of their particular loss function; see Diebold et al. (1998) and Granger and Pesaran (2000a,b). In order to compute the PIT in equation (1), one needs to assume a particular distribution function for $g_t(y_t|Y_{t-1})$. Simple tests of independence and uniformity, such as, the Kolmogorov-Smirnov test suffer from the problems described in the introduction. Alternatively, González-Rivera and Sun (2015) propose using autocontours to evaluate the PITs.

Define G-ACR$_{k,\alpha_i}$ as the set of points in the plane (u_t, u_{t-k}) such that the square with $\sqrt{\alpha_i}$-side contains α_i% of observations, i.e.,

$$\text{G-ACR}_{k,\alpha_i} = \{B(u_t, u_{t-k}) \subset \Re^2 | 0 \le u_t \le \sqrt{\alpha_i} \quad \text{and} \quad 0 \le u_{t-k} \le \sqrt{\alpha_i}, s.t. : u_t \times u_{t-k} \le \alpha_i\}. \tag{3}$$

Define also the following indicator series I_t^{k,α_i}:

$$I_t^{k,\alpha_i} = \mathbf{1}((u_t, u_{t-k}) \in \text{G-ACR}_{k,\alpha_i}) = \mathbf{1}(0 \le u_t \le \sqrt{\alpha_i}, 0 \le u_{t-k} \le \sqrt{\alpha_i}). \tag{4}$$

If $g_t(y_t|Y_{t-1})$ is a consistent estimator of $f_t(y_t|Y_{t-1})$, then I_t^{k,α_i} is an asymptotically Bernoulli MA process whose order depends on k. The sample proportion of PIT pairs (u_t, u_{t-k}) within the G-ACR$_{k,\alpha_i}$ cube is given by

$$\widehat{\alpha}_i = \frac{\sum\limits_{t=k+1}^{T} I_t^{k,\alpha_i}}{T-k}. \tag{5}$$

Consider the statistic t_{k,α_i}, given by

$$t_{k,\alpha_i} = \frac{\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)}{\sigma_{\alpha_i}}, \tag{6}$$

where $\sigma^2_{\alpha_i} = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$. González-Rivera and Sun (2015) show that under the null hypothesis in (2) the t_{k,α_i} statistics in (6) is asymptotically standard Normal distributed.

The t-statistic in (6) is constructed for a single fixed autocontour, α_i, and a single fixed lag, k. However, it can be generalized to a set of lags and a fixed autocontour or to several autocontours with a fixed lag. In the first case, for a fixed autocontour α_i, define $L_{\alpha_i} = (\ell_{1,\alpha_i}, ..., \ell_{K,\alpha_i})'$ which is a $K \times 1$ stacked vector with element $\ell_{k,\alpha_i} = \sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)$. Under H_0' in (2), $L'_{\alpha_i}\Lambda^{-1}_{\alpha_i}L_{\alpha_i}$ is asymptotically χ^2_K distributed, where a typical element of the asymptotic covariance matrix, Λ_{α_i}, is given by:

$$\lambda_{j,k} = \begin{cases} \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2}), & j = k, \\ 4\alpha_i^{3/2}(1-\alpha_i^{1/2}), & j \ne k. \end{cases}$$

Alternatively, for a fixed lag k, define $C_k = (c_{k,1}, ..., c_{k,C})'$ which is a $C \times 1$ stacked vector with element $c_{k,i} = \sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)$. Once more, under H_0' in (2), $C_k' \Omega_k^{-1} C_k$ has asymptotically a χ^2_C distribution, where a typical element of the asymptotic covariance matrix, Ω_k, is given by:

$$\omega_{i,j} = \begin{cases} \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2}), & i = j, \\ \alpha_i(1-\alpha_j) + 2\alpha_i\alpha_j^{1/2}(1-\alpha_j^{1/2}), & i < j, \\ \alpha_j(1-\alpha_i) + 2\alpha_j\alpha_i^{1/2}(1-\alpha_i^{1/2}), & i > j. \end{cases}$$

If the researcher is interested in partial aspects of the densities, such as, a particular collection of quantiles, it is more informative to examine the L_{α_i} statistic, which incorporates information for all desired k lags. On the other hand, if he is interested in the whole distribution, C_k collects information on all C autocontours desired, given a fixed lag k.

The tests described above are based on a given known predictive density $g_t(y_t|\Omega_{t-1})$. However, in practice, the parameters associated with the moments of this density need to be estimated. González-Rivera and Sun (2015) analyze the effects of parameter estimation on the asymptotic distribution of t_{k,α_i} and, consequently, on L_{α_i} and C_k, and conclude that the corresponding adjustments to the asymptotic variance are model dependent, and consequently, difficult to calculate analytically. So as to overcome this drawback, they propose a fully parametric bootstrap procedure to approximate the asymptotic variance based on obtaining random extractions from the known error predictive density assumed under the null hypothesis.

The G-ACR tests described above can be implemented both in-sample and out-of-sample. González-Rivera and Sun (2015) show that when testing the out-of-sample specification, the importance of parameter uncertainty will depend on both the forecasting scheme and the size of the estimation sample (T) relative to the forecast sample (H). Implementing the tests to out-of-sample forecast densities, the parameter uncertainty will distort their sizes as long as the proportion of the out-of-sample and in-sample sizes, H and T, respectively, is large. However, under the assumption of $\sqrt{T}$-consistent estimators, if $T \longrightarrow \infty$, $H \longrightarrow \infty$ and $H/T \longrightarrow 0$ as $T \longrightarrow \infty$, parameter uncertainty is asymptotic negligible, and no adjustment is needed for the test.

Finally, note that, if any of the G-ACR tests described above rejects the null hypothesis, there is not any indication about whether the rejection is due to an inadequate assumption about the error distribution or because the dynamics of the model are misspecified. González-Rivera and Sun (2015) point out that the G-ACR tests are more powerful for detecting departures from the distributional assumption than for detecting misspecified dynamics.

3. In-sample bootstrap BG-ACR tests

In this section, we propose a modification of the G-ACR test which allows testing for the specification of the conditional moments without making any particular assumption on the conditional distribution. We also justify heuristically the asymptotic distribution of the corresponding statistics and carry out Monte Carlo experiments to establish the finite sample performance of the new proposed tests.

3.1. Bootstrap predictive densities

Consider the following parametric model for the series of interest, y_t, $t = 1, ..., T$,

$$y_t = \mu_t + \sigma_t \varepsilon_t, \tag{7}$$

where μ_t and σ_t^2 are the conditional mean and variance of y_t, which are specified as parametric functions of Y_{t-1}. Finally, ε_t is a strict white noise process with distribution F_ε, such that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$. The parameters governing the conditional mean and variance need to be restricted to guarantee stationarity and the conditions required for their estimator to be consistent and asymptotically Normal. Asymptotic Normality of the parameter estimator is a requirement for the bootstrap to be asymptotically valid for the estimation of its sample distribution; see, for example, Hall and Yao (2003). Note that the asymptotic Normality of the estimator usually also depends on the distribution of the errors which should also be accordingly restricted.

A particular specification of (7) is following the popular AR(1)-GARCH(1,1) model which will be considered in this paper to illustrate the proposed tests

$$\begin{aligned} y_t &= \mu + \phi y_{t-1} + a_t, \\ a_t &= \varepsilon_t \sigma_t, \\ \sigma_t^2 &= \omega + \alpha a_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{8}$$

where $|\phi| < 1$, $\alpha + \beta < 1$, $\omega > 0$ and $\alpha, \beta \geq 0$. These assumptions are required to guarantee the stationarity of y_t and the positiveness of the conditional variance.

In this paper, we consider the Gaussian Quasi-Maximum Likelihood (QML) estimator of the parameters of the AR(1)-GARCH(1,1) model in (8) obtained by maximizing the Gaussian likelihood. Francq and Zakoian (2004) prove the strong consistency and asymptotic normality of the QML estimator of the ARMA-GARCH model under finite fourth order moment of the observed series.

It is worth noting that the procedure proposed in this paper to obtain bootstrap in-sample conditional densities and the consequent BG-ACR statistics to evaluate them, can be applied to any other parametric specifications of the conditional mean and variance as far as a consistent and asymptotically Normal estimator of the parameters is available; see, for example, Mika and Saikkonen (2011) who prove the strong consistency and asymptotic normality of the Gaussian QML estimator allowing both the conditional mean and the conditional variance to be nonlinear.

Next, we describe the bootstrap algorithm proposed to obtain in-sample one-step-ahead bootstrap densities of y_t in the context of the AR(1)-GARCH(1,1) model in (8). The algorithm is based on the residual bootstrap algorithms of Pascual et al. (2004, 2006) for linear ARMA models and GARCH models, respectively.

In-sample bootstrap algorithm

Step 1 Obtain the residuals

Estimate the parameters of model in (8) by a two-step QML estimator: $\widehat{\mu}$, $\widehat{\phi}$, $\widehat{\omega}$, $\widehat{\alpha}$ and $\widehat{\beta}$. Obtain the residuals $\widehat{\varepsilon}_t = \frac{\widehat{a}_t}{\widehat{\sigma}_t}$, $t = 3, ..., T$, where

$$\widehat{a}_t = y_t - \widehat{\mu} - \widehat{\phi} y_{t-1} \tag{9}$$

and

$$\widehat{\sigma}_t^2 = \widehat{\omega} + \widehat{\alpha}\widehat{a}_{t-1}^2 + \widehat{\beta}\widehat{\sigma}_{t-1}^2, \tag{10}$$

with $\widehat{a}_2 = y_2 - \widehat{\mu} - \widehat{\phi} y_1$ and $\widehat{\sigma}_2^2 = \widehat{\omega}/(1 - \widehat{\alpha} - \widehat{\beta})$. Denote by $F_{\widehat{\varepsilon}}$ the empirical distribution of the centered and scaled residuals.

Step 2 Bootstrap replicates of parameter estimates

For $t = 3, ..., T$, obtain recursively a bootstrap replicate of y_t that mimics the dynamic dependence of the original series as follows

$$\sigma_t^{*2(b)} = \widehat{\omega} + \widehat{\alpha} a_{t-1}^{*2(b)} + \widehat{\beta} \sigma_{t-1}^{*2(b)}, \tag{11}$$

$$a_t^{*(b)} = \varepsilon_t^{*(b)} \sigma_t^{*(b)},$$

$$y_t^{*(b)} = \widehat{\mu} + \widehat{\phi} y_{t-1}^{*(b)} + a_t^{*(b)}, \tag{12}$$

where $a_2^{*(b)} = \widehat{a}_2$, $\sigma_2^{*2(b)} = \widehat{\sigma}_2^2$, $y_2^{*(b)} = y_2$ and $\varepsilon_t^{*(b)}$ are random extractions with replacement from $F_{\widehat{\varepsilon}}$. Estimate the parameters by QML using $\left\{ y_t^{*(b)} \right\}_{t=3}^{T}$, obtaining $\widehat{\mu}^{*(b)}$, $\widehat{\phi}^{*(b)}$, $\widehat{\omega}^{*(b)}$, $\widehat{\alpha}^{*(b)}$ and $\widehat{\beta}^{*(b)}$.

Step 3 Obtain in-sample bootstrap one-step-ahead predictive densities

For $t = 3, ..., T$, obtain in-sample one-step-ahead estimates of volatilities and observations as follows:

$$\sigma_t^{**2(b)} = \widehat{\omega}^{*(b)} + \widehat{\alpha}^{*(b)} (y_{t-1} - \widehat{\mu}^{*(b)} - \widehat{\phi}^{*(b)} y_{t-2})^2 + \widehat{\beta}^{*(b)} \sigma_{t-1}^{**2(b)}, \tag{13}$$

$$y_t^{**(b)} = \widehat{\mu}^{*(b)} + \widehat{\phi}^{*(b)} y_{t-1} + \sigma_t^{**(b)} \varepsilon_t^{*(b)}, \tag{14}$$

where $\sigma_2^{**2(b)} = \widehat{\omega}^{*(b)} / (1 - \widehat{\alpha}^{*(b)} - \widehat{\beta}^{*(b)})$ and $\varepsilon_t^{*(b)}$ are random extractions with replacement from $F_{\widehat{\varepsilon}}$.

Step 4 Repeat steps 2 and 3 for $b = 1, ..., B^{(1)}$.

Note that in step 2, we obtain replicates of y_t^* which are not conditional on $\{y_1, ..., y_{t-1}\}$. In (11), σ_t^{*2} depends on a_{t-1}^{*2} while in (12) y_t^* depends on y_{t-1}^*. Therefore, independent replicates of the process are generated to estimate the parameters and to obtain an estimate of their sample distribution. However, in step 3, the bootstrap replicates, σ_t^{**2} and y_t^{**}, in (13) and (14), are obtained incorporating the parameter uncertainty through the bootstrap estimates of the parameters but always conditional on $\{y_1, ..., y_{t-1}\}$. In this way, at each moment of time, $t = 3, ..., T$, the above algorithm generates $B^{(1)}$ bootstrap replicates of y_t conditional on Y_{t-1}, which incorporate the parameter uncertainty and do not rely on any specific assumption about the distribution of ε_t. In order to decide the number of bootstrap replicates needed to obtain an appropriate estimate of the predictive density, one can implement the procedure proposed by Andrews and Buchinsky (2000). Note that the number of bootstrap replicates could be larger when dealing with non-linear GARCH errors than when the model is linear.

In-sample PITs can be easily computed as follows

$$u_t = \frac{1}{B^{(1)}} \sum_{b=1}^{B^{(1)}} \mathbf{1}(y_t^{**(b)} < y_t). \tag{15}$$

The corresponding indicators, I_t^{k,α_i}, and sample proportions, $\widehat{\alpha}_i$, can be computed as in (4) and (5), respectively. Finally, the t_{k,α_i}, L_{α_i} and C_k statistics can be calculated as explained above.

In order to illustrate how the proposed procedure works, we have generated a time series of size T=5000 from the following homoscedastic AR(1) model:

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{16}$$

where $\phi = (0.5, 0.95)$ and ε_t is i.i.d. with either N(0,1) or centered and standardized Student-5 and $\chi^2_{(5)}$ distributions. In each case, an AR(1) model is fitted to the artificial series with the parameters estimated by QML. Then, the in-sample PITs are computed both assuming normal errors as in González-Rivera and Sun (2015) and implementing the bootstrap algorithm described above based on $B^{(1)} = 999$ replicates; see Pascual et al. (2004, 2006) for the same number of replicates and Horváth et al. (2004) for $B^{(1)} = 1499$. Figure 1 plots the autocontours for α_i=0.2 and 0.8 together with the pairs (u_t, u_{t-1}) for the model AR(1) with ϕ=0.5 and $\varepsilon_t \sim N(0,1)$ (first row); ϕ=0.5 and ε_t ∼Student-5 (second row); ϕ=0.5 and $\varepsilon_t \sim \chi^2_{(5)}$ (third row); and ϕ=0.95 and $\varepsilon_t \sim \chi^2_{(5)}$ (fourth row). First of all, note that when the PITs are computed using the bootstrap densities (first column of Figure 1), they are uniformly distributed on the surface regardless of the true error distribution of the underlying DGP. Therefore, they suggest that the fitted AR(1) model is adequate. However, when the PITs are computed as in the G-ACR procedure (second column of Figure 1), they are not uniformly distributed unless the errors are Gaussian. In this case, the model is rejected but there is not indication about whether it is rejected because of the specification of the conditional mean or because of the assumed error distribution.

Consider now the following three DGPs:

$$y_t = 0.3y_{t-1} + 0.6y_{t-2} + \varepsilon_t. \tag{17}$$

$$y_t = \begin{cases} 0.5y_{t-1} + \varepsilon_t, & \text{for } t < T/2. \\ 1 + 0.5y_{t-1} + \varepsilon_t, & \text{for } t \geq T/2. \end{cases} \tag{18}$$

$$\begin{aligned} y_t &= 0.5y_{t-1} + \varepsilon_t\sigma_t. \\ \sigma_t^2 &= 0.05 + 0.5\varepsilon_{t-1}^2\sigma_{t-1}^2 + 0.45\sigma_{t-1}^2, \end{aligned} \tag{19}$$

with ε_t being an independent white noise with either N(0,1) or centered and standardized Student-5 or $\chi^2_{(5)}$ distributions. As above, an AR(1) model is fitted to each of the simulated series and its parameters estimated by QML. Then the PITs are computed assuming Normal errors and using the bootstrap procedure. Figure 2 plots the autocontours for α_i=0.2 and 0.8 together with the pairs (u_t, u_{t-1}) when the DGP is the AR(2) model in (17) with $\chi^2_{(5)}$ errors (first row); the AR(1) model with structural break in the mean in (18) with $\varepsilon_t \sim \chi^2_{(5)}$ (second row); the GARCH model in (19) with Normal errors (third row); and the GARCH model in (19) with $\chi^2_{(5)}$ errors (fourth row). We can observe that, when the PITs are based on the bootstrap densities (first column of Figure 2), they suggest the source of the misspecification. In the first row, when the AR(1) model is fitted to the AR(2) series, we observe a linear relation between the PITs. In the second row, when the DGP is the AR(1) model with a break in the mean, the PITs do not show any particular linearity or non-linearity but they are concentrated on the top-right corner of the plot. Finally, when the DGP is the AR(1)-GARCH(1,1) model, we observe a non-linear relation between the PITs. Furthermore, in this last case, the autocontour plots are very similar regardless of the error distribution of the DGP. Comparing the bootstrap-based PITs with those obtained using the normal densities (second column of Figure 2), the rejection of the fitted models is also clear although there is not an obvious indication for its source.

The asymptotic distributions of the t_{k,α_i}, L_{α_i} and C_k statistics depend on the asymptotic validity of the residual bootstrap algorithm described above. The asymptotic validity of the residual bootstrap procedure when implemented to obtain predictive densities in the context of linear ARMA models, has been established by Pascual et al. (2004). However, as far as we know, there is not a formal proof of the validity of the algorithm to construct predictive densities in the context of nonlinear GARCH models. In order to show that the algorithm is asymptotically valid, one needs first to show that the bootstrap procedure in step 2, generates asymptotically valid estimates of the model parameters. When implemented in GARCH models, Hidalgo and Zaffaroni (2007) show the first order validity of $\widehat{\theta}^* = (\widehat{\omega}^*, \widehat{\alpha}^*, \widehat{\beta}^*)$ for an ARCH(∞) process characterized by a particular decay in the ARCH parameters.[3] If the bootstrap procedure were asymptotically valid for the estimation of the parameters, using the arguments in Pascual et al. (2004) and Reeves (2005), one can establish its validity for the predictive densities and consequently, the distribution of $\widehat{\alpha}_i$ should be as in (6) with the asymptotic variance corrected to take into account the parameter uncertainty.[4]

Following the suggestion of González-Rivera and Sun (2015), the variance of $\widehat{\alpha}_i$ is approximated using a bootstrap procedure. $B^{(2)}$ bootstrap replicates, $\{y_t^{*(b)}\}_{t=1}^{T}$ are generated as in (12) and $\widehat{\alpha}_i^{*(b)}$ is obtained using the bootstrap series as if they were the original series. The bootstrap variance of $\widehat{\alpha}_i$ is given by

$$\sigma_{\alpha_i}^{*2} = \frac{1}{B^{(2)}-1} \sum_{b=1}^{B^{(2)}} \left(\widehat{\alpha}_i^{*(b)} - \frac{1}{B^{(2)}} \sum_{b=1}^{B^{(2)}} \widehat{\alpha}_i^{*(b)} \right)^2, \tag{20}$$

and the corresponding t-statistic is

$$t_{\alpha_i}^* = \frac{(\widehat{\alpha}_i - \alpha_i)}{\sigma_{\alpha_i}^*}, \tag{21}$$

which asymptotically has a N(0,1) distribution. In this paper, results are based on $B^{(2)}$=500 bootstrap replicates to compute $\sigma_{\alpha_i}^*$; see González-Rivera and Sun (2015). Note that the number of replicates needed to estimate standard errors is smaller than that needed to estimate intervals; see Efron (1987).

Obviously, the variances and covariances of the portmanteau statistics can also be computed using the same arguments. In particular, a typical element of the covariance

[3]Shimizu (2010, 2013, 2014) prove the consistency of the bootstrap QML estimators in the context of an AR(1)-ARCH(1) model. However, the residual bootstrap considered by Shimizu (2010, 2013, 2014) is not exactly the same as that considered in this paper. All the trajectories share the same estimated conditional mean and variance when generating bootstrap replicates to estimate the parameters. It is important to point out that Corradi and Iglesias (2008) cast some doubts on the asymptotic validity of the residual bootstrap described in step 2. Alternatively, they show that a block bootstrap based on resampling the likelihood as proposed by Gonçalves and White (2004) is asymptotically valid. Therefore, in step 2 of the algorithm described above, one can consider using the block bootstrap instead of the residual bootstrap.

[4]Monte Carlo results on the size distortions of the t-statistic when the asymptotic variance is computed as in (6) are available upon request.

matrix of L_{α_i}, say $\lambda^*_{j,k}$, is obtained as follows:

$$\lambda^*_{j,k} = \begin{cases} \sigma^{2*}_{\alpha_i}, & \text{if } j = k, \\ \frac{1}{B^{(2)}-1} \sum_{b=1}^{B^{(2)}} \left(\widehat{\alpha}^{*(b)}_{j,i} - \frac{1}{B^{(2)}} \sum_{b=1}^{B^{(2)}} \widehat{\alpha}^{*(b)}_{j,i} \right) \left(\widehat{\alpha}^{*(b)}_{k,i} - \frac{1}{B^{(2)}} \sum_{b=1}^{B^{(2)}} \widehat{\alpha}^{*(b)}_{k,i} \right), & \text{if } j \neq k, \end{cases} \tag{22}$$

Similarly, a typical element of the covariance matrix of C_k, say $\omega^*_{i,j}$, is obtained as follows:

$$\omega^*_{i,j} = \begin{cases} \sigma^{2*}_{\alpha_i}, & \text{if } i = j, \\ \frac{1}{B^{(2)}-1} \sum_{b=1}^{B^{(2)}} \left(\widehat{\alpha}^{*(b)}_{k,i} - \frac{1}{B^{(2)}} \sum_{b=1}^{B^{(2)}} \widehat{\alpha}^{*(b)}_{k,i} \right) \left(\widehat{\alpha}^{*(b)}_{k,j} - \frac{1}{B^{(2)}} \sum_{b=1}^{B^{(2)}} \widehat{\alpha}^{*(b)}_{k,j} \right), & \text{if } i \neq j. \end{cases} \tag{23}$$

3.2. Monte Carlo experiments

In this section, we perform Monte Carlo simulations to assess the finite sample properties of the proposed statistics. For the size assessment, the DPG is a linear AR(1). We consider a model far from the non-stationary region and another one near the non-stationary region with different error distributions. For the power assessment, we consider linear and non-linear alternatives. The number of Monte Carlo replicates is $R = 1000$ and the sample size $T = 50$, 100, 300, 1000 and 5000. The number of bootstrap replicates is $B^{(1)} = 1000$, except if $T = 5000$, when we use $B^{(1)} = 2000$. Finally, the number of bootstrap replicates used to compute the variance of $\widehat{\alpha}_i$, L_{α_i} and C_k is $B^{(2)} = 500$.

3.2.1. Studying the size

To investigate the size properties of the tests, we consider as DGP the AR(1) in equation (16). For each Monte Carlo replicate, we compute the proportions $\widehat{\alpha}_i$, for $k = 1, ..., 5$, and their bootstrap variances. Then, we compute the Monte Carlo averages and standard deviations of $\widehat{\alpha}_i$, together with the averages of the bootstrap standard deviations and the percentage of rejections of the null hypothesis when the nominal size of the test is 5%. Tables 1 and 2 report the Monte Carlo results for k=1 when $\phi = 0.5$ and the error is Gaussian and $\phi = 0.95$ and the errors are $\chi^2_{(5)}$, respectively. First of all, we observe that even for the smallest sample size of $T = 50$, the Monte Carlo averages of $\widehat{\alpha}_i$ are rather close to α_i and that the average of the bootstrap standard deviations is a good approximation to the Monte Carlo standard deviation of $\widehat{\alpha}_i$ for moderate sample sizes. However, note that for relatively small sample sizes, the bootstrap standard deviations tend to overestimate the empirical standard deviations of $\widehat{\alpha}_i$, mainly for the largest quantiles. Consequently, the size of the t_{1,α_i} statistic is smaller than the nominal. As the sample size increases, the percentage of rejections gets rather close to the 5% nominal level. The conclusions are quite similar for the close-to-unit-root model with $\chi^2_{(5)}$ errors.

We also analyze the finite sample performance of the two portmanteau tests. Table 3 reports the Monte Carlo percentage of rejections of $L^5_{\alpha_i}$ (adding up the information of the first five lags) and of C_1 (adding information of the thirteen quantiles previously considered) for the same two models considered in Tables 1 and 2. Looking at the results of $L^5_{\alpha_i}$, we observe that, regardless of the DGP considered, the Monte Carlo percentage of rejections is very close to the nominal size with a tendency to overreject for the largest quantiles. On the other hand, the results of C_1 show that it rejects less than the nominal size when the sample size is not large enough. However, if the sample size is large, the empirical size is larger than the nominal.

3.2.2. Studying the power

With the purpose of studying the finite sample power of the tests, we generate replicates by the three models in equations (17), (18) and (19) and fit an AR(1) model. Under the null hypothesis, we test the correct specification of the AR(1) model without drift. The DGP in (17) allows investigating the power against departures from the independence hypothesis, while the DGP in (18) deals with the power against breaks in the conditional mean. Finally, the DGP in (19) permits to analyze power when the second moment is misspecified.

Tables 4 to 6 report the power results of t_{1,α_i}, for each of the three DGPs. Consider first the results reported in Tables 4 when the DGP is the AR(2) model. We observe that t_{1,α_i} has high power for the intermediate autocontours around the 10%-60% levels even when the sample size is moderately small, that is, $T = 100$. As the sample size increases the power of t_{1,α_i} approaches one. Regarding the portmanteau tests, and in particular $L^5_{\alpha_i}$, we observe that its power is similar to the power of the t_{1,α_i} test, but its rejection rates are higher; see Panel A of Table 7. Furthermore, C_1 also shows high power, approaching one, even when $T = 300$. With respect to the DGP in (18), when corresponding to a break in the conditional mean, Table 5 shows that the power is also higher for the intermediate autocontours when the sample sizes are small. A remark to make is that, in this case, the t_{1,α_i} test is more powerful than the corresponding portmanteau tests of Panel B in Table 7. Finally, regarding the DGP in (19) when the series are generated by an AR(1)-GARCH(1,1) model, we observe in Tables 6 and 7 that the power of the t_{1,α_i} and $L^5_{\alpha_i}$ tests is higher in the extreme autocontours in comparison to the power reported by the intermediate autocontours, approaching one for $T = 5000$. The C_1 statistic also provides power close to one, but only in large sample sizes. Overall, the results suggest that larger sample sizes are needed to discriminate between the model under the null hypothesis and the GARCH model in (19).

4. Out-of-sample one-step-ahead bootstrap BG-ACR tests

In this section, we extend the procedures and tests described above to one-step-ahead out-of-sample densities. Note that the in-sample bootstrap algorithm can be also applied to obtain bootstrap replicates of y_{T+1} by implementing equations (13) and (14) for $T+1$. Then, the corresponding PIT, $u_{T+1|T} = u_{T+1}$ and the indicator $I^{k,\alpha_i}_{T+1|T} = I^{k,\alpha_i}_{T+1}$ can be computed as in (15). In order to compute the proportion, it is necessary to obtain H one-step-ahead bootstrap forecast densities. If the parameters are not reestimated each time a new observation is available, then the in-sample algorithm can be implemented as described in section 3 with step 3 modified as follows:

Step 3'. Obtain bootstrap one-step-ahead out-of-sample forecast densities

For $h = 1, ..., H$, obtain out-of-sample one-step-ahead estimates of volatilities and observations as follows:

$$\sigma^{**2(b)}_{T+h|T+h-1} = \widehat{\omega}^{*(b)} + \widehat{\alpha}^{*(b)}(y_{T+h-1} - \widehat{\mu}^{*(b)} - \widehat{\phi}^{*(b)} y_{T+h-2})^2 + \widehat{\beta}^{*(b)} \sigma^{**2(b)}_{T+h-1|T+h-2}, \quad (24)$$
$$y^{**(b)}_{T+h|T+h-1} = \widehat{\mu}^{*(b)} + \widehat{\phi}^{*(b)} y_{T+h-1} + \sigma^{**(b)}_{T+h|T+h-1} \varepsilon^{*(b)}_{T+h},$$

where $\sigma^{**2(b)}_{T+1|T} = \sigma^{**2(b)}_{T+1}$ in (13) and $\varepsilon^{*(b)}_{T+1}$ are random extractions with replacement from $F_{\widehat{\varepsilon}}$.

At each moment $T+h$, $h = 1, ..., H$, we compute the out-sample one-step-ahead PITs as follows

$$u_{T+h|T+h-1} = \frac{1}{B^{(1)}} \sum_{b=1}^{B^{(1)}} \mathbf{1}(y^{**(b)}_{T+h|T+h-1} < y_{T+h}).$$

Using $\{u_{T+h|T+h-1}\}^H_{h=1}$, we compute the corresponding indicators, I^{k,α_i}_{T+h}, and the proportion

$$\widehat{\alpha}_i = \frac{\sum\limits_{h=k+1}^{H} I_{T+h}^{k,\alpha_i}}{H-k}.$$

Finally, the t-statistic is given by

$$t_{k,\alpha_i} = \frac{\sqrt{H-k}(\widehat{\alpha}_i - \alpha_i)}{\sigma_{\alpha_i}},$$

where $\sigma^2_{\alpha_i}$ is defined as (6). Note that $\sigma^2_{\alpha_i}$ can be estimated either as in expression (6) or by bootstrapping. As mentioned above, when testing the in-sample specification, ignoring parameter uncertainty may cause severe distortions in the size of the tests. However, when testing the out-of-sample specification, the importance of parameter uncertainty decreases as far $H/T \to 0$ when $T \to \infty$. Therefore, if H is small relative to T, one can compute the variance, $\sigma^2_{\alpha_i}$, by using the asymptotic expression.

As an illustration of the out-of-sample performance of the tests, $R = 1000$ replicates are generated by the AR(1) model in expression (16) with $\phi = 0.95$ and $\varepsilon_t \sim N(0,1)$. The model is estimated by OLS using T=50, 100, 300, 1000 and 5000 observations and H=50 and 500 out-of-sample one-step ahead densities and their corresponding PITs are obtained using the bootstrap procedure. The variance of $\widehat{\alpha}_i$ and the covariances in Λ_{α_i} and Ω_k are computed by bootstrapping.[5] Table 8 reports the Monte Carlo averages and standard deviations of $\widehat{\alpha}_i$ for k=1, together with the averages of the bootstrap standard deviations and the percentage of rejections of the null hypothesis for different autocontours when the nominal size of the test is fixed at 5%. Table 9 reports the size of the corresponding $L^5_{\alpha_i}$ and C_1 test statistics. Table 8 shows that the size of the t-test when computed for out-of-sample one-step-ahead densities is close to the nominal as far as T is relatively large and H/T is small. Obviously, increasing H improves the size properties of the test as far as the ratio H/T is still small; see the size results of the L_{α_i} and C_k tests which are reported in Table 9 for H=500.

Finally, we study the finite sample power of the out-of-sample tests. With this purpose, R=1000 replicates are generated from the AR(2) model in (17) and the AR(1)-GARCH(1,1) model in (19). Under the null hypothesis, we consider an AR(1)

[5]Results based on the asymptotic expression of the variances and covariances are very similar when H=50 and T=1000 (H/T=0.05) or T=5000 (H/T=0.01). When H=500, the results are similar if T=5000 (H/T=0.1). As mentioned above, in these cases, the parameter uncertainty is irrelevant. These results are available upon request.

process without drift. Table 10 reports the power of the t_{1,α_i} test when the DGP is the AR(2) model and H=500. In this case, the power increases when the information is accumulated either over several lags or over several quantiles; see the powers reported in Panel A of Table 12. The results corresponding to the AR(1)-GARCH(1,1) model are reported in Table 11 for H=500. In this case, we can observe that the power of the t_{1,α_i} test is very low except when α_i=0.01. Furthermore, Panel B of Table 12 shows that the power does not increase when accumulating information over different autocontours or over different lags. Accumulating information in this way seems to be of no help when dealing with non-linear misspecifications.

5. Empirical application: modelling VIX

There is an increasing interest in modeling and forecasting the daily forward-looking market volatility index (VIX) from the Chicago Board Options Exchange (CBOE); see, for example, Whaley (2000, 2009), Engle and Gallo (2006), Fernandes et al. (2014), Diebold and Yilmaz (2015a) and Hassler et al. (2016). The VIX was originally computed as the weighted average of the implied volatilities from eight at-the-money call and put options of the S&P100 index with an average time to maturity of 30 days. In 2003, the VIX was entirely revised by changing the reference index to the S&P500 index, taking into account a wide range of strike prices for the same time to maturity and freeing its calculation from any specific option pricing model; see Whaley (2009) for a history of the VIX and Fernandes et al. (2014) for a detailed description of the VIX calculation. The VIX is important since it is a barometer of the overall market sentiment; see Whaley (2000, 2009) and Diebold and Yilmaz (2015b) who define it as a fear index. Furthermore, it reflects both the stock market uncertainty and the expected premium from selling stock market variance in a swap contract. Finally, there is an active market on VIX derivatives. The number of VIX futures contracts traded increased dramatically from about 1 million in 2007 to about 24 million in 2012 with the largest growth occurring after 2009, likely caused by the recent financial crisis; see, for example, Park (2016) and Song and Xiu (2016) for recent references on pricing VIX derivatives. The recent development of volatility-based derivative products generates an interest on predictive densities of volatility; see, for example, Corradi et al. (2009) who propose a feasible model free estimator of the conditional predictive density of integrated volatility based on subsampling. In the context of VIX, Konstantinidi and Skiadopoulos (2011) implement the bootstrap procedure of Pascual et al. (2004) to obtain forecast intervals for the VIX that are then used in a trading strategy. Konstantinidi et al.

(2008) and Psaradellis and Sermpinis (2016) also compare several specifications of the VIX for trading purposes. It is commonly accepted that the VIX display long-memory; see, for example, Bandi and Perron (2006), Konstantinidi et al. (2008), Shimotsu (2012), Fernandes et al. (2014) and Hassler et al. (2016). Consequently, several authors propose variants of the simple and easy-to-estimate approximate long-memory HAR model of Corsi (2009) to represent and predict the VIX; see Fernandes et al. (2014), Caporin et al. (2016) and Psaradellis and Sermpinis (2016). The HAR model is given by

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_5 \bar{y}_{t-1:5} + \phi_{10} \bar{y}_{t-1:10} + \phi_{22} \bar{y}_{t-1:22} + \phi_{66} \bar{y}_{t-1:66} + \varepsilon_t, \tag{25}$$

where $\bar{y}_{t:i} = i^{-1} \sum_{j=0}^{i-1} y_{t-j}$ and ε_t is an independent white noise sequence.

In this section, we analyze the series of daily log-VIX index observed from January 2, 1990 to January 15, 2003 with a total of 5807 observations; see Fernandes et al. (2014) for an empirical analysis of the same series. Descriptive statistics of the full sample are reported in Table 13. We can observe that the skewness and kurtosis are not significantly different from the assumed values under Normality when using the correction proposed by Premaratne and Bera (2016). However, the Jarque-Bera test clearly rejects the Normality of log-VIX. Panel (a) of Figure 3 plots the time series of the log-VIX. With respect to the temporal dependence, panels (b) and (c) of Figure 3 plot the correlograms of log-VIX and its squares, respectively. The comparison of the correlations of log-VIX and $(\text{log-VIX})^2$ suggests the presence of conditional heterocedasticity given that the correlations of squares are larger than those of the levels; see the values of the sample autocorrelations reported in Table 13. Fernandes et al. (2014) show that the null hypothesis of a unit-root is clearly rejected. Yet, they find strong evidence of long-memory.[6] The pure HAR model in equation (25) is fitted to the full sample with the estimated parameters reported in Table 14. Several residual diagnostics are reported in Table 13. We can observe that the distribution of the residuals is clearly non-Normal. Furthermore, the presence of conditional heterocedasticity is also more evident than when looking at the original log-VIX series. In-sample bootstrap conditional densities are computed as described in Section 3. Figure 4 plots kernel estimates of the bootstrap densities at different moments of the sample period. We can observe that not only the location of these densities changes

[6] Note that the unit-root tests carried out by Fernandes et al. (2014) do not take into account the presence of conditional heteroscedasticity. These tests should be modified as proposed by Choi (2015).

over time. Although, in general, there is a right skewness of the distribution, this skewness is more pronounced in some particular moments. Furthermore, we can also observe changes in the variance of the log-VIX. After computing the PITs, they are plotted in Figure 5, where we can observe that the PITs are not uniformly distributed. There is a concentration of PITs in the left and right top corners, suggesting that conditional heteroscedasticity has not been modeled when computing the conditional densities for log-VIX. For comparison purposes, we also plot in Figure 5, panel (b), the PITs computed by the procedure of González-Rivera and Sun (2015), where the errors are assumed to be Gaussian. We observe a concentration of points in the middle, suggesting that the HAR model is misspecified if Gaussian errors are assumed. However, as mentioned above, there is not indication of the source of misspecification.

In order to test the null hypothesis about the adequacy of the HAR model to represent the conditional densities of the log-VIX, Panel A of Table 15 reports sample proportions, $\widehat{\alpha}_i$, and the in-sample tests t_{1,α_i}, $L^5_{\alpha_i}$ and C^{13}_1. We observe that most of the autocontours are rejected by the t_{1,α_i} and $L^5_{\alpha_i}$ test statistics. The C^{13}_1 test, which is computed adding information of all autocontours, rejects H_0 at 1% of significance. Therefore, as suggested in Figure 5, the basic HAR model is not appropriate to model the conditional densities of the daily log-VIX. Panel B of Table 15 also reports the corresponding tests obtained by the procedure of González-Rivera and Sun (2015), assuming Gaussian errors. We can see that the null is rejected by almost all the autocontours, with the statistics being much larger than when they are computed using the BG-ACR tests.

Based on the information of the tests and autocontours, we incorporate conditional heteroscedasticity and estimate the following HAR-GARCH model:

$$\begin{aligned} y_t = \phi_0 + \phi_1 y_{t-1}\bar{y}_{t-1:5} + \phi_{10}\bar{y}_{t-10:5} + \phi_{22}\bar{y}_{t-1:22} + \phi_{66}\bar{y}_{t-1:66} + \sigma_t\varepsilon_t, \\ \sigma^2_t = \omega + \alpha\varepsilon^2_{t-1}\sigma^2_{t-1} + \beta\sigma^2_{t-1}. \end{aligned} \tag{26}$$

Table 14 reports the estimation results. We can observe that the estimates of the parameters of the conditional mean are very similar. Although the standard errors are different, the conclusions on their significance is the same as in the homoscedastic model. Furthermore, the parameters of the conditional variance equation are significative with values similar to those encountered when the GARCH model is fitted to financial returns with $\widehat{\alpha}$ being rather small and $\widehat{\alpha} + \widehat{\beta}$ very close to one. The residual statistics are reported in Table 13, where we can observe that the sample autocorrelations of the squares of log-VIX are no longer significant. Finally, Table 15, panels C and D, reports the results for the BG-ACR and G-ACR tests, respectively. We can observe that when implementing the G-ACR, the HAR-GARCH model with Gaussian errors is still clearly rejected. On the other hand, the number of rejections when implementing the BG-ACR tests is much smaller than before. In particular, when looking at the results for the $L^5_{\alpha_i}$, the HAR-GARCH model is only rejected for autocontours 0.3, 0.5 and 0.6, while the basic HAR model is rejected for eight out-of thirteen quintiles. Therefore, including the conditional heteroscedasticity leads to a better specification if the purpose is to obtain accurate predictive densities in-sample.

6. Conclusions

In this paper, we propose an extension of the G-ACR test of González-Rivera and Sun (2015) for dynamic specification of a density model (in-sample tests) and for evaluation of forecast densities (out-of-sample tests). Our contribution lies on computing the PITs from a bootstrapped conditional density so that no assumption on the functional form of the density is needed. Furthermore, the bootstrap procedure allows for the direct incorporation of parameter uncertainty. The proposed approach is particularly useful to evaluate forecast densities when the error distribution is unknown. Our proposed tests have size close to the nominal and are powerful for detecting departures from the assumed conditional density. To illustrate the usefulness of our approach, we extend the analysis of Fernandes et al. (2014) by evaluating the adequacy of conditional densities of the VIX daily market volatility index when computed fitting the HAR model. Our results suggest that conditional heteroscedasticity should be taken into account for an adequate construction of the conditional densities of the log-VIX.

In our research agenda, there are two direct extensions of the proposed BG-ACR tests that we plan to study. First, the extension of the proposed test to multi-step predictive densities is of great interest; see, for example, Jordà and Marcellino (2010), Staszewska-Bystrova (2011), Wolf and Wunderli (2015) and Jordá et al. (2014) for multistep forecasts based on bootstrap. Note that the functional form of multi-step predictive densities could be unknown or difficult to obtain even in cases where the one-step-ahead conditional density is known. Diebold et al. (1998) propose to partition the series of PITs into groups for which the iid uniformity is expected if the forecast densities were indeed correct. Analyzing this extension is left for future research.

Second, given that the tests considered in this paper are based on the information contained in the vector of PITs which is condensed into an indicator, the tests proposed can be extended into a multivariate framework using the multivariate bootstrap procedures of Fresoli et al. (2015) and Fresoli and Ruiz (2016) for VARMA and multivariate GARCH models, respectively. It is also important to note that in a multivariate context, the PITs with respect to a multivariate conditional density are not longer independent and uniform even if the model is correctly specified; see, for example, Chen and Hong (2014). In the context of multivariate GARCH models, Bai and Chen (2008) propose evaluating the distribution by using the PITs of each individual component. However, this test may miss important information on the joint distribution and, in particular, may fail to detect misspecification in the joint dynamics.

Finally, the residual bootstrap implemented in this paper to obtain one-step-ahead predictive densities can be modified in several directions. First, one can extended it to cope with lag-order uncertainty of the ARMA lags by implementing the procedures of Kilian (1998) and Alonso et al. (2004, 2006). Another alternative is substituting the basic residual bootstrap implemented in this paper to obtain the sample distribution of the parameters by the subsampling procedure proposed by Hall and Yao (2003). Alternatively, one can implement the block bootstrap based on resampling the likelihood proposed by Corradi and Iglesias (2008). Although, we do not expect the results to change qualitatively, the asymptotic validity of the bootstrap can be easier to prove in the case of GARCH errors.

References

Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014). Central Bank Macroeconomic Forecasting during the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York experiences. *Journal of Business & Economic Statistics 32*(4), 483–500.

Alonso, A. M., D. Peña, and J. Romo (2004). Introducing model uncertainty in time series bootstrap. *Statistica Sinica 14*, 155–174.

Alonso, A. M., D. Peña, and J. Romo (2006). Introducing model uncertainty by moving blocks bootstrap. *Statistical Papers 47*, 167–179.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Andrews, D. W. K. and M. Buchinsky (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica 68*(1), 23–51.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics 85*(3), 531–549.

Bai, J. and Z. Chen (2008). Testing multivariate distributions in GARCH models. *Journal of Econometrics 143*(1), 19–36.

Bandi, F. M. and B. Perron (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics 4*(4), 636–670.

Baumeister, C. and L. Kilian (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics 30*(2), 326–336.

Berkowitz, J. (2001). Testing density forecasts with applications to risk management. *Journal of Business & Economic Statistics 19*(4), 465–474.

Britton, E., P. Fisher, and J. Whitley (1998). The inflation report projections: understanding the fan chart. *Bank of England. Quarterly Bulletin 38*(1), 30.

Caporin, M., E. Rossi, and P. Santucci de Magistris (2016). Volatility jumps and their economic determinants. *Journal of Financial Econometrics 1*(14), 28–29.

Chen, B. and Y. Hong (2014). A unified approach to validating univariate and multivariate conditional distribution models in time series. *Journal of Econometrics 178*, 22–44.

Chen, X. and Y. Fan (2004). Evaluating density forecasts via the copula approach. *Finance Research Letters 1*(1), 74–84.

Choi, I. (2015). *Almost All About Unit Roots. Foundations, Developments and Applications.* New York: Cambridge University Press.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Clements, M. P., A. B. Galvão, and J. H. Kim (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance 15*(4), 729–750.

Corradi, V., W. Distaso, and N. R. Swanson (2009). Predictive density estimators for daily volatility based on the use of realized measures. *Journal of Econometrics 150*(2), 119–138.

Corradi, V. and E. M. Iglesias (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics 144*(2), 500–510.

Corradi, V. and N. Swanson (2006). Predictive density evaluation. In G. Elliot, C. Granger, and A. Timmerman (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 197–284. London: Elsevier.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 2*(7), 174–196.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–882.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics 20*(1), 134–144.

Diebold, F. X. and K. Yilmaz (2015a). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press, USA.

Diebold, F. X. and K. Yilmaz (2015b). Trans-Atlantic equity volatility connectedness: U.S. and European financial institutions. *Journal of Financial Econometrics 14*(1), 81–127.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association 82*(397), 171–185.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1), 3–27.

Fair, R. C. (1980). Estimating the expected predictive accuracy of econometric models. *International Economic Review 21*(2), 355–378.

Fernandes, M., M. C. Medeiros, and M. Scharth (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance 40*, 1–10.

Francq, C. and J.-M. Zakoian (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10*(4), 605–637.

Fresoli, D. and E. Ruiz (2016). The uncertainty of conditional returns, volatilities and correlations in DCC models. *Computational Statistics & Data Analysis 100*, 170–185.

Fresoli, D., E. Ruiz, and L. Pascual (2015). Bootstrap multi-step forecasts of non-gaussian VAR models. *International Journal of Forecasting 31*(3), 834–848.

Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin (2003). Forecast uncertainties in macroeconomic modeling: An application to the U.K. economy. *Journal of the American Statistical Association 98*(464), 829–838.

Giordani, P. and M. Villani (2010). Forecasting macroeconomic time series with locally adaptive signal extraction. *International Journal of Forecasting 26*(2), 312–325.

Gonçalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics 119*(1), 199–219.

González-Rivera, G., Z. Senyuz, and E. Yoldas (2011). Autocontours: Dynamic specification testing. *Journal of Business & Economic Statistics 29*(1), 186–200.

González-Rivera, G. and Y. Sun (2015). Generalized autocontours: evaluation of multivariate density models. *International Journal of Forecasting 31*(3), 799–814.

González-Rivera, G. and E. Yoldas (2012). Autocontour-based evaluation of multivariate predictive densities. *International Journal of Forecasting 28*(2), 328–342.

Granger, C. W. J. and M. H. Pesaran (2000a). A decision theoretic approach to forecast evaluation. In W. S. Chan, W. K. Li, and H. Tong (Eds.), *Statistics and Finance: An Interface*, Chapter 15, pp. 261–278. London: Imperial College Press.

Granger, C. W. J. and M. H. Pesaran (2000b). Economic and statistical measures of forecast accuracy. *Journal of Forecasting 19*, 537–560.

Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH models with heavy–tailed errors. *Econometrica 71*(1), 285–317.

Hallam, M. and J. Olmo (2013). Semiparametric density forecasts of daily financial returns from intraday data. *Journal of Financial Econometrics 12*(2), 408–432.

Hallam, M. and J. Olmo (2014). Forecasting daily return densities from intraday data: A multifractal approach. *International Journal of Forecasting 30*(4), 863–881.

Hassler, U., P. M. M. Rodrigues, and A. Rubia (2016). Quantile regression for long memory testing: A case of realized volatility. *Journal of Financial Econometrics*, forthcoming.

Hidalgo, J. and P. Zaffaroni (2007). A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics 141*(2), 973–1013.

Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies 18*(1), 37–84.

Horváth, L., P. Kokoszka, and G. Teyssière (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation 74*(7), 469–485.

Jordá, Ò., M. Knüppel, and M. Marcellino (2014). Empirical simultaneous prediction regions for path-forecasts. *International Journal of Forecasting 29*, 456–468.

Jordà, Ò. and M. Marcellino (2010). Path-forecast evaluation. *Journal of Applied Econometrics 25*(4), 635–662.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics 25*(4), 621–634.

Kilian, L. (1998). Accounting for lag order uncertainty in autoregressions: The endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis 19*(5), 531–548.

Konstantinidi, E. and G. Skiadopoulos (2011). Are VIX futures prices predictable? An empirical investigation. *International Journal of Forecasting 27*(2), 543–560.

Konstantinidi, E., G. Skiadopoulos, and E. Tzagkaraki (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking & Finance 32*(11), 2401–2411.

Lam, J. P. and M. Veall (2002). Bootstrap prediction intervals for single period regression forecasts. *International Journal of Forecasting 18*(1), 125–130.

Maheu, J. M. and T. H. McCurdy (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics 160*(1), 69–76.

Mika, M. and P. Saikkonen (2011). Parameter estimation in nonlinear AR–GARCH models. *Econometric Theory 27*(6), 1236–1278.

Nieto, M. R. and E. Ruiz (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting 32*(2), 475–501.

Park, Y.-H. (2016). The effects of asymmetric volatility and jumps on the pricing of VIX derivatives. *Journal of Econometrics 192*(1), 313–328.

Pascual, L., J. Romo, and E. Ruiz (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis 25*(4), 449–465.

Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns, and volatilities in GARCH models. *Computational Statistics and Data Analysis 50*(9), 2146–2166.

Patton, A. and A. Timmermann (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association 102*(480), 1172–1184.

Premaratne, G. and A. Bera (2016). Adjusting the tests for skewness and kurtosis for distributional misspecifications. *Communications in Statistics-Simulation and Computation*, forthcoming.

Psaradellis, I. and G. Sermpinis (2016). Modelling and trading the US implied volatility indices. evidence from the VIX, VXN and VXD indices. *International Journal of Forecasting 32*(4), 1268–1283.

Ravazzolo, F. and P. Rothman (2016). Oil-price density forecasts of U.S. GDP. *Studies in Nonlinear Dynamics & Econometrics 20*(4), 441–453.

Reeves, J. J. (2005). Bootstrap prediction intervals for ARCH models. *International Journal of Forecasting 21*(2), 237–248.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*(3), 470–472.

Shimizu, K. (2010). *Bootstrapping Stationary ARMA-GARCH Models.* Springer, Vieweg.

Shimizu, K. (2013). The bootstrap does not always work for heteroscedastic models. *Statistics & Risk Modeling 30*(3), 189–204.

Shimizu, K. (2014). Bootstrapping the nonparametric ARCH regression model. *Statistics & Probability Letters 87*, 61–69.

Shimotsu, K. (2012). Exact local Whittle estimation of fractionally cointegrated systems. *Journal of Econometrics 169*(2), 266–278.

Song, Z. and D. Xiu (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics 190*(1), 176–196.

Soyer, E. and R. M. Hogarth (2012). The illusion of predictability: How regression statistics mislead experts. *International Journal of Forecasting 28*(3), 695–711.

Staszewska-Bystrova, A. (2011). Bootstrap prediction bands for forecast paths from vector autoregressive models. *Journal of Forecasting 30*(8), 721–735.

Tay, A. S. and K. F. Wallis (2000). Density forecasting: A survey. *Journal of Forecasting 19*(4), 235–254.

Tsay, R. S. (1992). Model checking via parametric bootstraps in time series analysis. *Applied Statistics 41*(1), 1–15.

Whaley, R. E. (2000). The investor fear gauge. *Journal of Portfolio Management 26*(3), 12–17.

Whaley, R. E. (2009). Understanding the VIX. *Journal of Portfolio Management 35*(3), 98–105.

Wolf, M. and D. Wunderli (2015). Bootstrap joint prediction regions. *Journal of Time Series Analysis 36*(3), 352–376.

Figures and tables

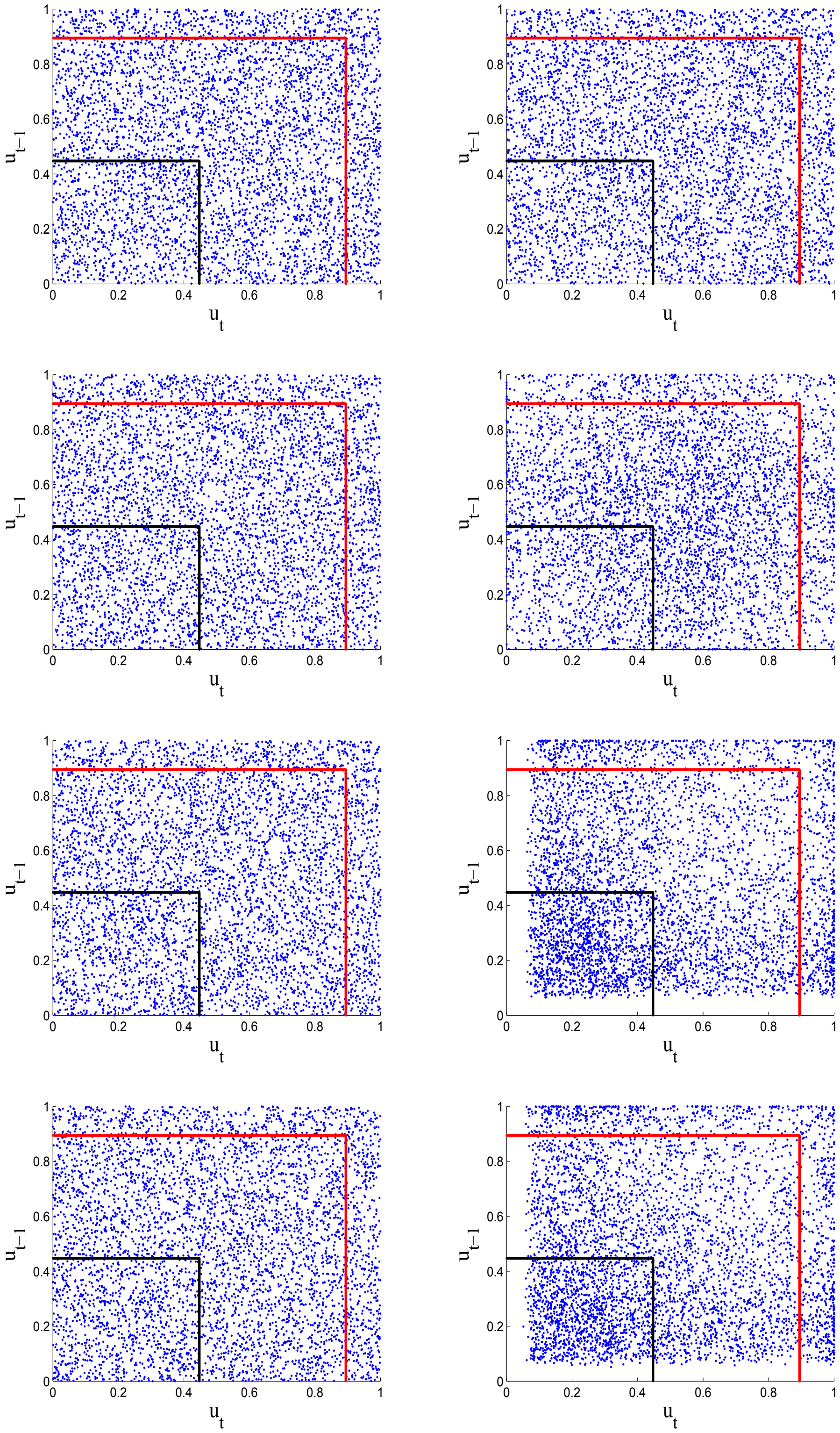

Figure 1: Univariate autocontours for the estimated AR(1) model with $T = 5000$. $ACR_{20\%,1}$ corresponds to the black box and the $ACR_{80\%,1}$ to the red box. The DGPs are the AR(1) model with: ϕ=0.5 and $\varepsilon_t \sim N(0,1)$ (first row); ϕ=0.5 and $\varepsilon_t \sim$ Student-5 (second row); ϕ=0.5 and $\varepsilon_t \sim \chi^2_{(5)}$ (third row); and ϕ=0.95 and $\varepsilon_t \sim \chi^2_{(5)}$ (fourth row). The PITs were computed using the bootstrap algorithm with $B^{(1)}$=1000 (first column), or assuming Gaussian errors (second column).

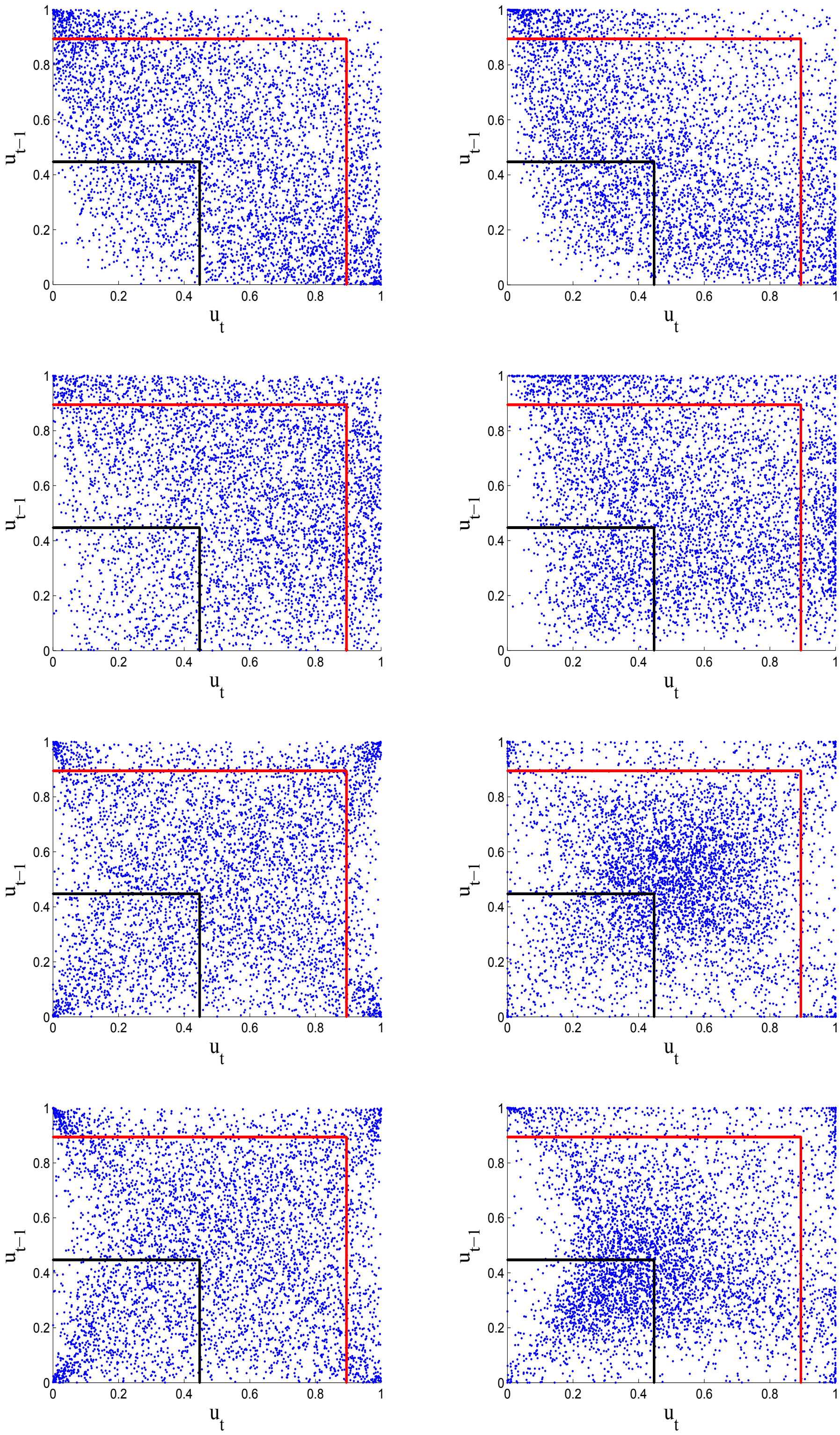

Figure 2: Univariate autocontours for estimated AR(1) model with T=5000. $ACR_{20\%,1}$ corresponds to the black box and the $ACR_{80\%,1}$ to the red box. The DGPs are: AR(2) with $\varepsilon_t \sim \chi^2_{(5)}$ (first row); AR(1) model with break in the mean with $\varepsilon_t \sim \chi^2_{(5)}$ (second row); AR(1)-GARCH(1,1) model with $\varepsilon_t \sim$ N(0,1) (third row); and AR(1)-GARCH(1,1) model with $\varepsilon_t \sim \chi^2_{(5)}$ (fourth row). The PITs were computed using the bootstrap algorithm with $B^{(1)}$=1000 (first column), or assuming Gaussian errors (second column).

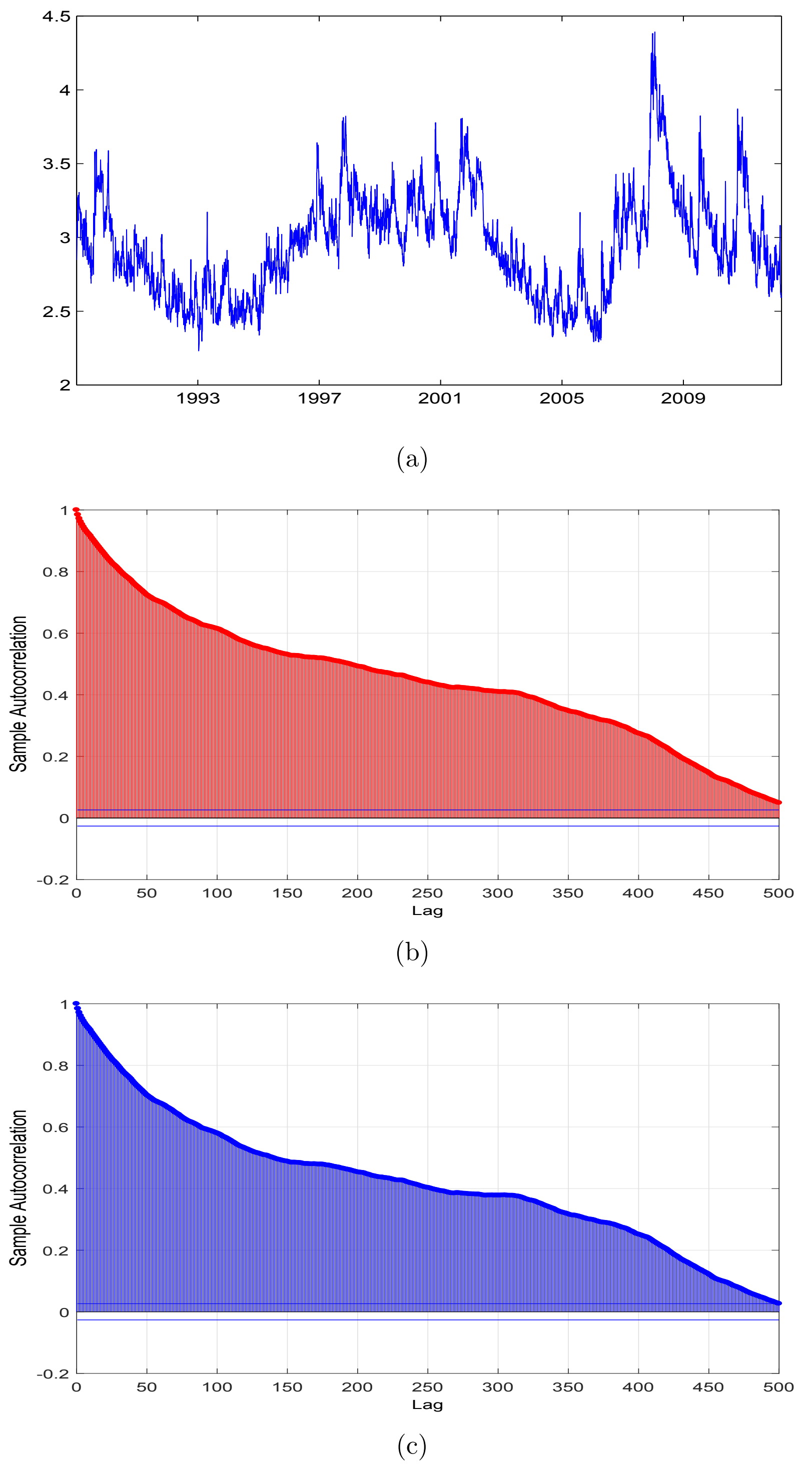

Figure 3: Daily log-VIX is plotted in (a). In (b) and (c) are plotted the sample autocorrelations of the levels and squares of the log-VIX, respectively.

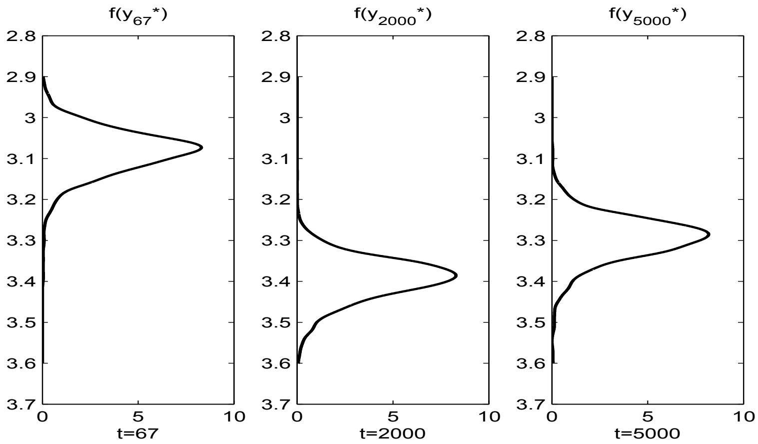

Figure 4: In-sample bootstrap one-step-ahead densities.

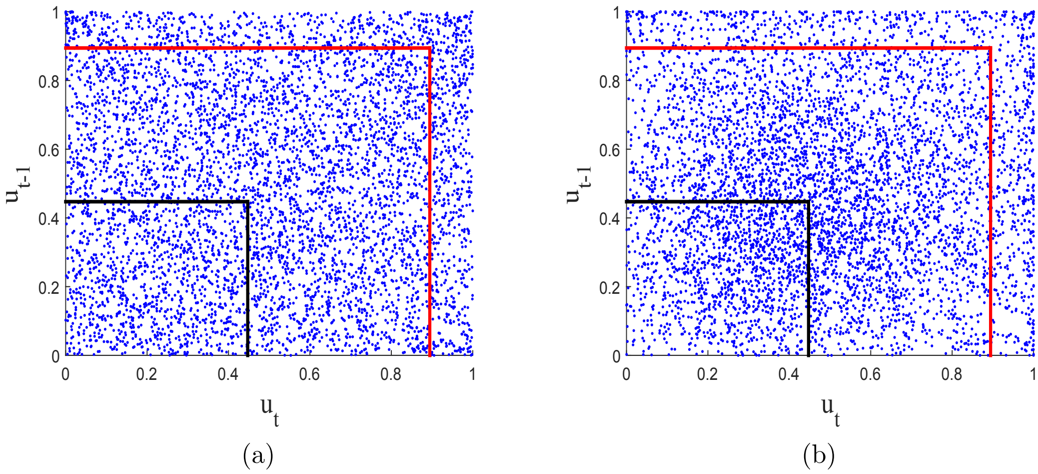

Figure 5: Univariate autocontours for the HAR model. In (a) the PITS are obtained with the bootstrap procedure described in Section 3 and in (b) they are obtained by the procedure of González-Rivera and Sun (2015) assuming Gaussian errors. $ACR_{20\%,1}$ corresponds to the black box and $ACR_{80\%,1}$ to the red box.

Table 1: Monte Carlo size results for t_{1,α_i}. The DGP is $y_t = 0.5y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim N(0,1)$ and the nominal size is 5%.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.009	0.049	0.102	0.204	0.309	0.405	0.507	0.606	0.706	0.803	0.903	0.950	0.986
	std	(0.014)	(0.033)	(0.045)	(0.066)	(0.082)	(0.088)	(0.089)	(0.087)	(0.080)	(0.068)	(0.051)	(0.037)	(0.024)
	$\bar{\sigma}^*_{\alpha_i}$	0.015	0.034	0.050	0.071	0.085	0.093	0.096	0.094	0.088	0.077	0.060	0.049	0.034
	size	0.058	0.031	0.027	0.026	0.035	0.023	0.027	0.021	0.026	0.019	0.017	0.007	0.024
100														
	$\widehat{\alpha}_i$	0.009	0.050	0.101	0.202	0.304	0.403	0.501	0.600	0.703	0.801	0.901	0.951	0.989
	std	(0.009)	(0.022)	(0.032)	(0.044)	(0.054)	(0.058)	(0.061)	(0.057)	(0.051)	(0.043)	(0.031)	(0.023)	(0.013)
	$\bar{\sigma}^*_{\alpha_i}$	0.010	0.023	0.033	0.048	0.057	0.062	0.064	0.062	0.057	0.049	0.037	0.028	0.018
	size	0.042	0.032	0.041	0.031	0.036	0.028	0.034	0.031	0.023	0.025	0.016	0.016	0.009
300														
	$\widehat{\alpha}_i$	0.009	0.050	0.100	0.201	0.300	0.400	0.500	0.599	0.699	0.799	0.899	0.949	0.988
	std	(0.005)	(0.012)	(0.017)	(0.025)	(0.029)	(0.031)	(0.032)	(0.031)	(0.028)	(0.023)	(0.016)	(0.011)	(0.006)
	$\bar{\sigma}^*_{\alpha_i}$	0.006	0.013	0.018	0.026	0.031	0.033	0.034	0.033	0.030	0.025	0.018	0.013	0.008
	size	0.019	0.036	0.039	0.035	0.033	0.028	0.031	0.030	0.031	0.031	0.014	0.018	0.026
1000														
	$\widehat{\alpha}_i$	0.010	0.050	0.100	0.200	0.299	0.399	0.500	0.599	0.699	0.799	0.898	0.949	0.988
	std	(0.003)	(0.007)	(0.010)	(0.014)	(0.017)	(0.017)	(0.018)	(0.017)	(0.015)	(0.012)	(0.008)	(0.006)	(0.003)
	$\bar{\sigma}^*_{\alpha_i}$	0.003	0.007	0.010	0.014	0.016	0.017	0.018	0.017	0.015	0.013	0.009	0.006	0.003
	size	0.050	0.049	0.048	0.046	0.049	0.049	0.044	0.050	0.049	0.041	0.038	0.038	0.065
5000														
	$\widehat{\alpha}_i$	0.010	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.799	0.900	0.950	0.989
	std	(0.001)	(0.003)	(0.004)	(0.006)	(0.007)	(0.008)	(0.008)	(0.007)	(0.006)	(0.005)	(0.004)	(0.002)	(0.001)
	$\bar{\sigma}^*_{\alpha_i}$	0.001	0.003	0.004	0.006	0.007	0.008	0.008	0.007	0.007	0.005	0.003	0.002	0.001
	size	0.044	0.070	0.052	0.046	0.048	0.054	0.052	0.055	0.039	0.057	0.052	0.044	0.127

Table 2: Monte Carlo size results for t_{1,α_i}. The DGP is $y_t = 0.95 y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \chi^2_{(5)}$ and the nominal size is 5%.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.015	0.058	0.109	0.209	0.307	0.407	0.504	0.603	0.705	0.804	0.901	0.950	0.984
	std	(0.022)	(0.048)	(0.067)	(0.089)	(0.098)	(0.102)	(0.097)	(0.094)	(0.081)	(0.064)	(0.043)	(0.034)	(0.023)
	$\bar{\sigma}^*_{\alpha_i}$	0.022	0.049	0.068	0.090	0.100	0.103	0.101	0.095	0.086	0.072	0.054	0.045	0.032
	size	0.061	0.046	0.026	0.022	0.015	0.023	0.016	0.025	0.013	0.012	0.001	0.004	0.009
100														
	$\widehat{\alpha}_i$	0.012	0.055	0.106	0.205	0.305	0.406	0.503	0.602	0.702	0.803	0.900	0.949	0.989
	std	(0.014)	(0.032)	(0.045)	(0.060)	(0.064)	(0.067)	(0.062)	(0.057)	(0.049)	(0.038)	(0.027)	(0.020)	(0.012)
	$\bar{\sigma}^*_{\alpha_i}$	0.015	0.033	0.046	0.060	0.066	0.067	0.065	0.060	0.053	0.043	0.032	0.025	0.018
	size	0.060	0.037	0.029	0.025	0.021	0.025	0.014	0.012	0.014	0.014	0.008	0.005	0.000
300														
	$\widehat{\alpha}_i$	0.011	0.052	0.102	0.202	0.303	0.402	0.502	0.601	0.701	0.800	0.899	0.949	0.988
	std	(0.007)	(0.017)	(0.024)	(0.030)	(0.033)	(0.032)	(0.032)	(0.028)	(0.024)	(0.018)	(0.013)	(0.009)	(0.006)
	$\bar{\sigma}^*_{\alpha_i}$	0.008	0.017	0.024	0.031	0.034	0.034	0.032	0.030	0.026	0.020	0.014	0.011	0.007
	size	0.044	0.036	0.033	0.039	0.034	0.022	0.032	0.024	0.031	0.017	0.026	0.018	0.011
1000														
	$\widehat{\alpha}_i$	0.011	0.051	0.101	0.201	0.301	0.401	0.501	0.600	0.700	0.800	0.899	0.949	0.988
	std	(0.004)	(0.009)	(0.012)	(0.016)	(0.017)	(0.017)	(0.016)	(0.015)	(0.012)	(0.009)	(0.006)	(0.004)	(0.003)
	$\bar{\sigma}^*_{\alpha_i}$	0.004	0.009	0.012	0.016	0.017	0.017	0.016	0.015	0.012	0.010	0.007	0.005	0.003
	size	0.054	0.048	0.046	0.051	0.039	0.040	0.042	0.048	0.045	0.034	0.037	0.043	0.101
5000														
	$\widehat{\alpha}_i$	0.010	0.050	0.101	0.200	0.300	0.400	0.500	0.600	0.700	0.799	0.900	0.950	0.989
	std	(0.002)	(0.004)	(0.005)	(0.007)	(0.008)	(0.007)	(0.007)	(0.006)	(0.005)	(0.004)	(0.002)	(0.002)	(0.001)
	$\bar{\sigma}^*_{\alpha_i}$	0.002	0.004	0.005	0.007	0.007	0.007	0.007	0.006	0.005	0.004	0.002	0.002	0.001
	size	0.049	0.063	0.055	0.046	0.054	0.039	0.049	0.046	0.051	0.044	0.051	0.056	0.162

Table 3: Monte Carlo size results for $L^5_{\alpha_i}$ and C_1 statistics. The DGPs are: $y_t = 0.5y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$ (Panel A) and $y_t = 0.95y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \chi^2_{(5)}$ (Panel B). The nominal size is 5%.

	$L^5_{0.01}$	$L^5_{0.05}$	$L^5_{0.1}$	$L^5_{0.2}$	$L^5_{0.3}$	$L^5_{0.4}$	$L^5_{0.5}$	$L^5_{0.6}$	$L^5_{0.7}$	$L^5_{0.8}$	$L^5_{0.9}$	$L^5_{0.95}$	$L^5_{0.99}$	C^{13}_1
T							**Panel A**							
50	0.078	0.066	0.055	0.052	0.047	0.034	0.049	0.044	0.068	0.089	0.089	0.137	0.076	0.025
100	0.048	0.065	0.056	0.047	0.040	0.042	0.051	0.056	0.055	0.058	0.074	0.108	0.049	0.023
300	0.057	0.048	0.053	0.044	0.041	0.040	0.040	0.051	0.056	0.048	0.067	0.091	0.073	0.029
1000	0.050	0.051	0.047	0.045	0.042	0.042	0.049	0.048	0.051	0.046	0.062	0.061	0.179	0.054
5000	0.062	0.060	0.051	0.042	0.042	0.046	0.061	0.043	0.044	0.054	0.060	0.052	0.101	0.092
							Panel B							
50	0.121	0.091	0.073	0.043	0.036	0.033	0.048	0.059	0.064	0.063	0.081	0.107	0.058	0.023
100	0.098	0.065	0.053	0.045	0.038	0.041	0.054	0.048	0.051	0.038	0.091	0.115	0.027	0.008
300	0.083	0.051	0.057	0.032	0.047	0.045	0.041	0.040	0.036	0.050	0.060	0.095	0.075	0.023
1000	0.070	0.053	0.053	0.058	0.056	0.055	0.047	0.051	0.053	0.051	0.058	0.070	0.193	0.060
5000	0.063	0.051	0.051	0.048	0.043	0.044	0.044	0.034	0.050	0.062	0.051	0.048	0.120	0.089

Table 4: Monte Carlo power results for t_{1,α_i}. The DGP is the AR(2) model in (17) with $\varepsilon_t \sim N(0,1)$. The nominal size is 5%.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.000	0.001	0.006	0.035	0.095	0.184	0.294	0.420	0.553	0.693	0.833	0.902	0.953
	std	(0.000)	(0.005)	(0.012)	(0.029)	(0.042)	(0.050)	(0.052)	(0.055)	(0.053)	(0.049)	(0.040)	(0.031)	(0.021)
	$\bar{\sigma}^*_{\alpha_i}$	0.017	0.041	0.060	0.085	0.100	0.109	0.112	0.109	0.100	0.085	0.063	0.049	0.032
	power	0.000	0.000	0.084	0.491	0.588	0.534	0.389	0.227	0.110	0.069	0.050	0.046	0.116
100														
	$\widehat{\alpha}_i$	0.000	0.002	0.008	0.044	0.111	0.205	0.320	0.446	0.577	0.712	0.848	0.916	0.970
	std	(0.001)	(0.004)	(0.010)	(0.022)	(0.030)	(0.035)	(0.038)	(0.037)	(0.035)	(0.032)	(0.025)	(0.020)	(0.014)
	$\bar{\sigma}^*_{\alpha_i}$	0.011	0.027	0.040	0.058	0.068	0.074	0.075	0.073	0.067	0.056	0.041	0.031	0.019
	power	0.000	0.268	0.830	0.947	0.961	0.925	0.831	0.636	0.354	0.195	0.080	0.075	0.060
300														
	$\widehat{\alpha}_i$	0.000	0.002	0.011	0.054	0.128	0.228	0.344	0.469	0.600	0.731	0.861	0.926	0.979
	std	(0.000)	(0.003)	(0.006)	(0.013)	(0.018)	(0.020)	(0.022)	(0.020)	(0.019)	(0.018)	(0.013)	(0.011)	(0.007)
	$\bar{\sigma}^*_{\alpha_i}$	0.006	0.014	0.021	0.029	0.034	0.037	0.038	0.036	0.033	0.027	0.019	0.014	0.008
	power	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.869	0.544	0.360	0.227
1000														
	$\widehat{\alpha}_i$	0.000	0.002	0.012	0.058	0.135	0.237	0.355	0.479	0.609	0.739	0.866	0.930	0.982
	std	(0.000)	(0.002)	(0.004)	(0.008)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.009)	(0.007)	(0.006)	(0.003)
	$\bar{\sigma}^*_{\alpha_i}$	0.003	0.007	0.011	0.015	0.017	0.018	0.018	0.018	0.016	0.013	0.009	0.006	0.003
	power	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.948	0.652
5000														
	$\widehat{\alpha}_i$	0.000	0.002	0.013	0.060	0.138	0.240	0.358	0.484	0.612	0.741	0.868	0.932	0.983
	std	(0.000)	(0.001)	(0.002)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)
	$\bar{\sigma}^*_{\alpha_i}$	0.001	0.003	0.005	0.006	0.007	0.008	0.008	0.007	0.007	0.005	0.004	0.002	0.001
	power	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999

Table 5: Monte Carlo power results for t_{1,α_i}. The DGP is the AR(1) model with break in the mean in (18) with $\varepsilon_t \sim N(0,1)$. The nominal size is 5%.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.001	0.014	0.040	0.103	0.176	0.264	0.359	0.462	0.577	0.701	0.830	0.897	0.948
	std	(0.005)	(0.016)	(0.026)	(0.041)	(0.049)	(0.055)	(0.059)	(0.059)	(0.060)	(0.052)	(0.045)	(0.038)	(0.027)
	$\bar{\sigma}^*_{\alpha_i}$	0.016	0.037	0.053	0.076	0.090	0.099	0.102	0.099	0.092	0.079	0.061	0.049	0.033
	power	0.000	0.000	0.041	0.105	0.131	0.144	0.137	0.145	0.142	0.090	0.098	0.080	0.163
100														
	$\widehat{\alpha}_i$	0.002	0.017	0.043	0.109	0.184	0.273	0.372	0.477	0.589	0.712	0.840	0.909	0.967
	std	(0.004)	(0.013)	(0.019)	(0.027)	(0.033)	(0.039)	(0.042)	(0.041)	(0.039)	(0.036)	(0.029)	(0.026)	(0.017)
	$\bar{\sigma}^*_{\alpha_i}$	0.010	0.024	0.035	0.050	0.059	0.064	0.065	0.063	0.058	0.050	0.037	0.029	0.018
	power	0.001	0.107	0.284	0.417	0.488	0.516	0.486	0.460	0.441	0.374	0.258	0.212	0.136
300														
	$\widehat{\alpha}_i$	0.002	0.018	0.046	0.114	0.193	0.283	0.381	0.485	0.599	0.719	0.849	0.917	0.976
	std	(0.003)	(0.007)	(0.011)	(0.016)	(0.020)	(0.022)	(0.024)	(0.025)	(0.024)	(0.021)	(0.018)	(0.014)	(0.008)
	$\bar{\sigma}^*_{\alpha_i}$	0.006	0.013	0.019	0.027	0.032	0.034	0.035	0.033	0.030	0.025	0.018	0.013	0.008
	power	0.000	0.785	0.927	0.985	0.989	0.994	0.995	0.994	0.986	0.969	0.874	0.716	0.421
1000														
	$\widehat{\alpha}_i$	0.002	0.019	0.047	0.116	0.196	0.287	0.385	0.490	0.604	0.724	0.851	0.920	0.979
	std	(0.002)	(0.004)	(0.006)	(0.009)	(0.011)	(0.012)	(0.013)	(0.013)	(0.013)	(0.012)	(0.009)	(0.008)	(0.004)
	$\bar{\sigma}^*_{\alpha_i}$	0.003	0.007	0.010	0.014	0.017	0.018	0.018	0.017	0.015	0.013	0.009	0.006	0.003
	power	0.824	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.911
5000														
	$\widehat{\alpha}_i$	0.003	0.019	0.047	0.116	0.197	0.287	0.387	0.491	0.604	0.724	0.853	0.921	0.980
	std	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)
	$\bar{\sigma}^*_{\alpha_i}$	0.001	0.003	0.004	0.006	0.007	0.008	0.008	0.007	0.007	0.005	0.004	0.002	0.001
	power	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6: Monte Carlo power results for t_{1,α_i}. The DGP is the AR(1)-GARCH(1,1) model in (19) with $\varepsilon_t \sim N(0,1)$. The nominal size is 5%.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.018	0.063	0.111	0.209	0.312	0.418	0.522	0.622	0.719	0.815	0.908	0.953	0.991
	std	(0.019)	(0.039)	(0.056)	(0.078)	(0.089)	(0.096)	(0.093)	(0.086)	(0.077)	(0.061)	(0.041)	(0.031)	(0.019)
	$\bar{\sigma}^*_{\alpha_i}$	0.014	0.035	0.054	0.079	0.094	0.103	0.104	0.100	0.090	0.075	0.057	0.047	0.033
	power	0.204	0.073	0.048	0.034	0.022	0.025	0.019	0.012	0.014	0.010	0.006	0.006	0.009
100														
	$\widehat{\alpha}_i$	0.021	0.066	0.112	0.207	0.308	0.413	0.517	0.620	0.718	0.816	0.909	0.953	0.990
	std	(0.014)	(0.029)	(0.043)	(0.061)	(0.071)	(0.075)	(0.072)	(0.064)	(0.050)	(0.038)	(0.026)	(0.018)	(0.011)
	$\bar{\sigma}^*_{\alpha_i}$	0.010	0.025	0.039	0.058	0.069	0.074	0.074	0.069	0.060	0.047	0.033	0.025	0.018
	power	0.321	0.132	0.067	0.043	0.051	0.035	0.028	0.027	0.017	0.015	0.009	0.004	0.002
300														
	$\widehat{\alpha}_i$	0.023	0.066	0.112	0.206	0.306	0.410	0.516	0.618	0.718	0.816	0.908	0.953	0.989
	std	(0.009)	(0.022)	(0.031)	(0.041)	(0.046)	(0.048)	(0.045)	(0.039)	(0.031)	(0.021)	(0.012)	(0.008)	(0.005)
	$\bar{\sigma}^*_{\alpha_i}$	0.006	0.015	0.024	0.035	0.042	0.045	0.044	0.041	0.034	0.025	0.015	0.011	0.007
	power	0.542	0.232	0.094	0.052	0.045	0.044	0.038	0.043	0.040	0.051	0.031	0.006	0.006
1000														
	$\widehat{\alpha}_i$	0.024	0.066	0.111	0.204	0.303	0.408	0.514	0.616	0.717	0.814	0.908	0.953	0.989
	std	(0.006)	(0.011)	(0.016)	(0.023)	(0.026)	(0.027)	(0.026)	(0.021)	(0.017)	(0.012)	(0.006)	(0.004)	(0.002)
	$\bar{\sigma}^*_{\alpha_i}$	0.003	0.008	0.013	0.020	0.024	0.026	0.025	0.023	0.019	0.013	0.007	0.005	0.003
	power	0.929	0.494	0.152	0.061	0.048	0.051	0.066	0.070	0.105	0.159	0.145	0.070	0.031
5000														
	$\widehat{\alpha}_i$	0.024	0.066	0.110	0.202	0.302	0.406	0.512	0.615	0.716	0.814	0.908	0.954	0.989
	std	(0.003)	(0.007)	(0.010)	(0.014)	(0.016)	(0.017)	(0.016)	(0.015)	(0.012)	(0.007)	(0.003)	(0.002)	(0.001)
	$\bar{\sigma}^*_{\alpha_i}$	0.001	0.004	0.006	0.010	0.011	0.012	0.012	0.011	0.009	0.006	0.003	0.002	0.001
	power	0.999	0.925	0.419	0.119	0.098	0.113	0.197	0.330	0.482	0.692	0.724	0.476	0.110

Table 7: Monte Carlo power results for $L^5_{\alpha_i}$ and C_1 statistics. The DGPs are: the AR(2) model in (17) (Panel A), the AR(1) model with break in the mean in (18) (Panel B) and the AR(1)-GARCH(1,1) in (19) (Panel C). The nominal size is 5%.

	$L^5_{0.01}$	$L^5_{0.05}$	$L^5_{0.1}$	$L^5_{0.2}$	$L^5_{0.3}$	$L^5_{0.4}$	$L^5_{0.5}$	$L^5_{0.6}$	$L^5_{0.7}$	$L^5_{0.8}$	$L^5_{0.9}$	$L^5_{0.95}$	$L^5_{0.99}$	C^{13}_1
T							**Panel A**							
50	0.009	0.057	0.249	0.643	0.774	0.796	0.793	0.720	0.582	0.417	0.281	0.310	0.228	0.058
100	0.027	0.299	0.723	0.939	0.975	0.976	0.968	0.934	0.854	0.678	0.475	0.331	0.179	0.441
300	0.343	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989	0.851	0.632	0.317	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.638	1.000
5000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000
							Panel B							
50	0.000	0.006	0.019	0.063	0.121	0.155	0.205	0.257	0.269	0.280	0.257	0.300	0.240	0.054
100	0.002	0.014	0.047	0.131	0.256	0.292	0.326	0.362	0.379	0.375	0.391	0.404	0.186	0.166
300	0.004	0.339	0.586	0.789	0.869	0.891	0.900	0.891	0.879	0.839	0.726	0.591	0.276	0.855
1000	0.743	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984	0.676	1.000
5000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
							Panel C							
50	0.177	0.093	0.064	0.054	0.041	0.040	0.045	0.031	0.055	0.085	0.122	0.180	0.052	0.059
100	0.301	0.104	0.076	0.065	0.051	0.050	0.055	0.057	0.056	0.095	0.207	0.279	0.061	0.107
300	0.589	0.175	0.071	0.064	0.061	0.063	0.074	0.084	0.088	0.143	0.282	0.473	0.314	0.381
1000	0.935	0.366	0.144	0.090	0.088	0.091	0.106	0.161	0.238	0.331	0.520	0.653	0.886	0.907
5000	0.999	0.875	0.345	0.166	0.154	0.187	0.332	0.557	0.770	0.890	0.940	0.941	0.972	1.000

Table 8: Monte Carlo size results for out-of-sample t_{1,α_i}. The DGP is $y_t = 0.95y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0,1)$. The nominal size is 5% and $H = 50$.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.015	0.061	0.112	0.213	0.313	0.409	0.508	0.602	0.699	0.794	0.891	0.935	0.968
	std	(0.033)	(0.064)	(0.085)	(0.117)	(0.133)	(0.146)	(0.152)	(0.150)	(0.142)	(0.123)	(0.095)	(0.072)	(0.050)
	$\bar{\sigma}^*_{\alpha_i}$	0.032	0.065	0.088	0.117	0.135	0.144	0.148	0.146	0.138	0.123	0.096	0.076	0.052
	size	0.045	0.058	0.058	0.064	0.071	0.080	0.072	0.084	0.083	0.057	0.053	0.052	0.075
100														
	$\widehat{\alpha}_i$	0.013	0.058	0.109	0.209	0.305	0.404	0.502	0.600	0.697	0.795	0.893	0.940	0.980
	std	(0.021)	(0.048)	(0.069)	(0.093)	(0.112)	(0.121)	(0.124)	(0.121)	(0.115)	(0.104)	(0.077)	(0.059)	(0.033)
	$\bar{\sigma}^*_{\alpha_i}$	0.022	0.050	0.071	0.098	0.114	0.123	0.127	0.126	0.118	0.104	0.079	0.060	0.036
	size	0.049	0.052	0.050	0.050	0.060	0.060	0.061	0.053	0.049	0.048	0.044	0.059	0.070
300														
	$\widehat{\alpha}_i$	0.011	0.054	0.105	0.209	0.310	0.408	0.510	0.608	0.705	0.804	0.899	0.949	0.986
	std	(0.016)	(0.041)	(0.055)	(0.082)	(0.100)	(0.109)	(0.111)	(0.109)	(0.103)	(0.088)	(0.067)	(0.049)	(0.026)
	$\bar{\sigma}^*_{\alpha_i}$	0.017	0.040	0.058	0.081	0.096	0.104	0.108	0.106	0.100	0.088	0.067	0.049	0.026
	size	0.038	0.048	0.034	0.055	0.063	0.058	0.066	0.053	0.050	0.038	0.040	0.050	0.056
1000														
	$\widehat{\alpha}_i$	0.011	0.051	0.103	0.205	0.304	0.404	0.501	0.600	0.700	0.800	0.897	0.949	0.987
	std	(0.016)	(0.037)	(0.055)	(0.075)	(0.090)	(0.097)	(0.100)	(0.100)	(0.094)	(0.083)	(0.063)	(0.046)	(0.023)
	$\bar{\sigma}^*_{\alpha_i}$	0.016	0.037	0.054	0.075	0.088	0.096	0.100	0.099	0.093	0.082	0.062	0.046	0.023
	size	0.070	0.045	0.049	0.045	0.055	0.049	0.051	0.054	0.046	0.053	0.041	0.041	0.049
5000														
	$\widehat{\alpha}_i$	0.010	0.050	0.099	0.202	0.300	0.398	0.502	0.603	0.701	0.800	0.896	0.947	0.987
	std	(0.014)	(0.035)	(0.051)	(0.072)	(0.085)	(0.093)	(0.097)	(0.097)	(0.090)	(0.078)	(0.058)	(0.043)	(0.023)
	$\bar{\sigma}^*_{\alpha_i}$	0.016	0.037	0.052	0.073	0.086	0.094	0.097	0.096	0.091	0.080	0.060	0.044	0.021
	size	0.050	0.043	0.043	0.036	0.048	0.051	0.053	0.057	0.044	0.052	0.025	0.032	0.048

Table 9: Monte Carlo size results for out-of-sample $L^5_{\alpha_i}$ and C_1 statistics. The DGP is $y_t = 0.95y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0,1)$. The nominal size is 5%, $H = 50$ (Panel A) and $H = 500$ (Panel B).

	$L^5_{0.01}$	$L^5_{0.05}$	$L^5_{0.1}$	$L^5_{0.2}$	$L^5_{0.3}$	$L^5_{0.4}$	$L^5_{0.5}$	$L^5_{0.6}$	$L^5_{0.7}$	$L^5_{0.8}$	$L^5_{0.9}$	$L^5_{0.95}$	$L^5_{0.99}$	C_1^{13}
T							**Panel A**							
50	0.116	0.113	0.096	0.083	0.070	0.072	0.090	0.091	0.108	0.108	0.121	0.117	0.147	0.086
100	0.100	0.075	0.084	0.050	0.039	0.051	0.062	0.080	0.107	0.105	0.116	0.136	0.094	0.078
300	0.120	0.080	0.068	0.059	0.066	0.073	0.069	0.070	0.077	0.091	0.118	0.139	0.087	0.071
1000	0.124	0.081	0.075	0.072	0.067	0.063	0.079	0.075	0.090	0.103	0.126	0.151	0.074	0.079
5000	0.093	0.077	0.054	0.058	0.053	0.062	0.063	0.062	0.073	0.079	0.116	0.161	0.090	0.064
							Panel B							
50	0.119	0.092	0.088	0.087	0.082	0.087	0.085	0.070	0.076	0.084	0.093	0.107	0.107	0.057
100	0.100	0.076	0.079	0.066	0.078	0.067	0.065	0.060	0.073	0.074	0.102	0.111	0.115	0.047
300	0.094	0.069	0.070	0.057	0.056	0.052	0.066	0.059	0.066	0.059	0.081	0.102	0.197	0.062
1000	0.074	0.063	0.059	0.055	0.060	0.059	0.065	0.056	0.075	0.068	0.081	0.090	0.164	0.065
5000	0.056	0.054	0.049	0.058	0.047	0.058	0.057	0.053	0.053	0.051	0.077	0.113	0.127	0.050

Table 10: Monte Carlo power results for out-of-sample t_{1,α_i}. The DGP is the AR(2) model in (17) with $\varepsilon_t \sim N(0,1)$. The nominal size is 5% and $H = 500$.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.000	0.002	0.008	0.038	0.100	0.193	0.305	0.430	0.559	0.692	0.829	0.898	0.948
	std	(0.001)	(0.003)	(0.008)	(0.021)	(0.038)	(0.059)	(0.079)	(0.095)	(0.102)	(0.098)	(0.083)	(0.068)	(0.049)
	$\bar{\sigma}^*_{\alpha_i}$	0.024	0.043	0.056	0.069	0.077	0.083	0.086	0.086	0.084	0.078	0.064	0.051	0.036
	power	0.000	0.030	0.450	0.764	0.809	0.745	0.653	0.526	0.418	0.327	0.234	0.212	0.211
100														
	$\widehat{\alpha}_i$	0.000	0.002	0.010	0.047	0.115	0.211	0.324	0.449	0.579	0.711	0.845	0.914	0.967
	std	(0.000)	(0.003)	(0.007)	(0.019)	(0.032)	(0.047)	(0.059)	(0.069)	(0.076)	(0.073)	(0.060)	(0.047)	(0.030)
	$\bar{\sigma}^*_{\alpha_i}$	0.012	0.026	0.037	0.049	0.056	0.061	0.064	0.064	0.063	0.057	0.045	0.035	0.021
	power	0.000	0.415	0.874	0.956	0.952	0.898	0.795	0.644	0.482	0.353	0.261	0.226	0.196
300														
	$\widehat{\alpha}_i$	0.000	0.002	0.011	0.055	0.129	0.229	0.344	0.469	0.599	0.730	0.859	0.925	0.978
	std	(0.000)	(0.002)	(0.006)	(0.015)	(0.023)	(0.033)	(0.040)	(0.046)	(0.049)	(0.047)	(0.038)	(0.028)	(0.015)
	$\bar{\sigma}^*_{\alpha_i}$	0.007	0.016	0.023	0.032	0.038	0.042	0.044	0.044	0.043	0.038	0.029	0.022	0.012
	power	0.000	0.999	1.000	1.000	1.000	0.991	0.951	0.832	0.665	0.444	0.310	0.244	0.194
1000														
	$\widehat{\alpha}_i$	0.000	0.002	0.012	0.058	0.134	0.237	0.355	0.482	0.610	0.739	0.866	0.930	0.982
	std	(0.000)	(0.002)	(0.006)	(0.012)	(0.019)	(0.026)	(0.033)	(0.038)	(0.039)	(0.036)	(0.029)	(0.022)	(0.011)
	$\bar{\sigma}^*_{\alpha_i}$	0.005	0.013	0.018	0.026	0.031	0.034	0.035	0.035	0.033	0.029	0.023	0.017	0.008
	power	0.149	1.000	1.000	1.000	1.000	1.000	0.985	0.898	0.736	0.521	0.349	0.258	0.205
5000														
	$\widehat{\alpha}_i$	0.000	0.002	0.012	0.059	0.139	0.239	0.358	0.484	0.614	0.741	0.869	0.933	0.985
	std	(0.000)	(0.002)	(0.005)	(0.012)	(0.018)	(0.023)	(0.028)	(0.033)	(0.033)	(0.030)	(0.025)	(0.018)	(0.009)
	$\bar{\sigma}^*_{\alpha_i}$	0.005	0.012	0.017	0.023	0.028	0.030	0.031	0.031	0.029	0.026	0.020	0.014	0.007
	power	0.813	1.000	1.000	1.000	1.000	1.000	1.000	0.961	0.802	0.608	0.385	0.246	0.187

Table 11: Monte Carlo power results for out-of-sample t_{1,α_i}. The DGP is the AR(1)-GARCH(1,1) model in (19) with $\varepsilon_t \sim N(0,1)$. The nominal size is 5% and $H = 500$.

T	α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
50														
	$\widehat{\alpha}_i$	0.030	0.070	0.115	0.213	0.323	0.432	0.537	0.630	0.718	0.800	0.875	0.912	0.943
	std	(0.022)	(0.036)	(0.047)	(0.063)	(0.082)	(0.098)	(0.109)	(0.114)	(0.110)	(0.100)	(0.083)	(0.069)	(0.054)
	$\bar{\sigma}^*_{\alpha_i}$	0.011	0.027	0.038	0.053	0.062	0.068	0.070	0.071	0.069	0.064	0.053	0.044	0.033
	power	0.388	0.183	0.119	0.090	0.118	0.148	0.198	0.219	0.243	0.248	0.193	0.228	0.289
100														
	$\widehat{\alpha}_i$	0.028	0.069	0.114	0.210	0.314	0.421	0.525	0.625	0.719	0.808	0.889	0.929	0.963
	std	(0.019)	(0.030)	(0.039)	(0.053)	(0.064)	(0.079)	(0.087)	(0.091)	(0.090)	(0.084)	(0.068)	(0.054)	(0.037)
	$\bar{\sigma}^*_{\alpha_i}$	0.009	0.021	0.032	0.045	0.052	0.057	0.058	0.057	0.054	0.049	0.040	0.031	0.020
	power	0.464	0.228	0.128	0.059	0.078	0.143	0.180	0.213	0.246	0.280	0.263	0.209	0.290
300														
	$\widehat{\alpha}_i$	0.026	0.067	0.113	0.207	0.308	0.413	0.519	0.619	0.718	0.811	0.899	0.943	0.977
	std	(0.015)	(0.025)	(0.032)	(0.040)	(0.048)	(0.058)	(0.066)	(0.069)	(0.069)	(0.065)	(0.052)	(0.040)	(0.024)
	$\bar{\sigma}^*_{\alpha_i}$	0.006	0.016	0.024	0.034	0.040	0.044	0.045	0.043	0.040	0.035	0.027	0.021	0.011
	power	0.551	0.281	0.135	0.066	0.065	0.120	0.169	0.214	0.265	0.305	0.314	0.300	0.271
1000														
	$\widehat{\alpha}_i$	0.025	0.067	0.112	0.205	0.305	0.410	0.515	0.617	0.716	0.813	0.905	0.950	0.985
	std	(0.012)	(0.019)	(0.024)	(0.030)	(0.035)	(0.041)	(0.046)	(0.049)	(0.051)	(0.049)	(0.040)	(0.031)	(0.017)
	$\bar{\sigma}^*_{\alpha_i}$	0.005	0.013	0.019	0.028	0.033	0.035	0.036	0.035	0.033	0.029	0.022	0.016	0.008
	power	0.587	0.317	0.151	0.059	0.073	0.097	0.127	0.198	0.238	0.286	0.298	0.304	0.206
5000														
	$\widehat{\alpha}_i$	0.024	0.066	0.110	0.202	0.303	0.408	0.513	0.617	0.716	0.814	0.908	0.954	0.989
	std	(0.012)	(0.018)	(0.022)	(0.026)	(0.031)	(0.036)	(0.041)	(0.045)	(0.047)	(0.045)	(0.036)	(0.027)	(0.013)
	$\bar{\sigma}^*_{\alpha_i}$	0.005	0.012	0.017	0.024	0.029	0.031	0.032	0.031	0.030	0.026	0.020	0.014	0.007
	power	0.632	0.330	0.148	0.065	0.066	0.087	0.133	0.189	0.244	0.302	0.320	0.316	0.124

Table 12: Monte Carlo size results for out-of-sample $L^5_{\alpha_i}$ and C_1 statistics. The DGPs are: the AR(2) model in (17) with $\varepsilon_t \sim N(0,1)$ and $H = 500$ (Panel A); and the AR(1)-GARCH(1,1) model with $\varepsilon_t \sim N(0,1)$ and $H = 500$ (Panel B). The nominal size is 5%.

	$L^5_{0.01}$	$L^5_{0.05}$	$L^5_{0.1}$	$L^5_{0.2}$	$L^5_{0.3}$	$L^5_{0.4}$	$L^5_{0.5}$	$L^5_{0.6}$	$L^5_{0.7}$	$L^5_{0.8}$	$L^5_{0.9}$	$L^5_{0.95}$	$L^5_{0.99}$	C^{13}_1
Panel A														
50	0.288	0.773	0.978	1.000	1.000	1.000	1.000	1.000	0.987	0.919	0.713	0.521	0.348	0.596
100	0.418	0.928	0.999	1.000	1.000	1.000	1.000	1.000	0.996	0.964	0.751	0.526	0.297	0.728
300	0.565	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.980	0.827	0.573	0.415	0.989
1000	0.691	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.986	0.832	0.602	0.385	1.000
5000	0.721	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.987	0.845	0.624	0.327	1.000
Panel B														
50	0.371	0.150	0.111	0.096	0.122	0.128	0.157	0.189	0.223	0.259	0.377	0.448	0.528	0.385
100	0.445	0.185	0.122	0.087	0.092	0.118	0.153	0.186	0.242	0.297	0.378	0.500	0.558	0.495
300	0.525	0.242	0.110	0.088	0.095	0.108	0.132	0.164	0.252	0.341	0.441	0.491	0.599	0.569
1000	0.569	0.238	0.126	0.077	0.071	0.078	0.097	0.170	0.239	0.360	0.479	0.500	0.503	0.573
5000	0.602	0.277	0.115	0.064	0.071	0.079	0.113	0.161	0.231	0.322	0.484	0.551	0.412	0.600

Table 13: Descriptive statistics

The first column corresponds to the summary statistics of the log-VIX series and the second and third to the residuals obtained with the HAR and HAR-GARCH models, respectively. ρ_k corresponds to the estimated sample autocorrelations of the log-VIX and ρ_k^2 to the sample correlations of its squares. k corresponds to the lag of the sample correlations. p-values are in parenthesis.

Sample statistics	Full sample	HAR model residuals	HAR-GARCH model residuals
Mean	2.953	0.000	0.001
Standard deviation	0.348	0.060	0.060
Skewness	0.539	0.915	0.946
	(0.000)	(0.000)	(0.000)
Kurtosis	3.288	7.481	7.523
	(0.000)	(0.000)	(0.000)
Jarque-Bera	300.683	5605.800	5749.000
	(0.000)	(0.001)	(0.001)
ρ_1	0.985	0.001	0.007
	(0.000)	(0.940)	(0.584)
ρ_{10}	0.916	0.052	0.044
	(0.000)	(0.016)	(0.057)
ρ_{100}	0.616	-0.002	0.014
	(0.000)	(0.000)	(0.000)
ρ_1^2	0.984	0.122	-0.003
	(0.000)	(0.000)	(0.830)
ρ_{10}^2	0.914	0.129	0.025
	(0.000)	(0.000)	(0.285)
ρ_{100}^2	0.582	-0.011	0.011
	(0.000)	(0.000)	(0.864)

Table 14: Estimation results for the log-VIX index. t-statistics in parenthesis.

	HAR	HAR-GARCH
	Mean equation	
	Estimate	Estimate
$\widehat{\phi}_0$	0.024	0.029
	(3.325)	(4.369)
$\widehat{\phi}_1$	0.873	0.883
	(64.422)	(65.136)
$\widehat{\phi}_2$	-0.002	-0.009
	(-0.058)	(-0.344)
$\widehat{\phi}_3$	0.133	0.108
	(4.510)	(3.605)
$\widehat{\phi}_4$	-0.030	-0.013
	(-1.556)	(-0.710)
$\widehat{\phi}_5$	0.016	0.022
	(2.070)	(2.799)
	Variance equation	
		Estimate
$\widehat{\omega}$		2.784e-05
		(11.057)
$\widehat{\alpha}$		0.088
		(14.384)
$\widehat{\beta}$		0.834
		(71.694)

Table 15: Testing the models in-sample

The sample period runs from January 2, 1990 to January 15, 2003, including altogether 5807 observations. Panels A and C shows the tests obtained by the bootstrap procedure described in Section 3 and panels B and D presents the tests obtained by the procedure of González-Rivera and Sun (2015), where the errors are assumed to be Gaussian and the variance of the tests is computed by bootstrap. *, **, *** indicate that H_0 is rejected at 10%, 5% and 1% levels of significance, respectively.

HAR model													
α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
Panel A													
$\widehat{\alpha}_i$	0.009	0.052	0.105	0.203	0.309	0.412	0.512	0.610	0.705	0.803	0.899	0.950	0.989
$\lvert t_{1,\alpha_i}\rvert$	$\lvert 0.51\rvert$	$\lvert 1.16\rvert$	$\lvert 1.71\rvert^{*}$	$\lvert 0.96\rvert$	$\lvert 2.79\rvert^{***}$	$\lvert 3.71\rvert^{***}$	$\lvert 4.08\rvert^{***}$	$\lvert 3.86\rvert^{***}$	$\lvert 2.31\rvert^{**}$	$\lvert 1.42\rvert$	$\lvert 0.61\rvert$	$\lvert 0.15\rvert$	$\lvert 0.97\rvert$
$L^5_{\alpha_i}$	5.85	3.03	6.24	5.27	11.67^{**}	18.51^{***}	21.978^{***}	16.84^{***}	9.92^{*}	4.83	11.34^{**}	12.66^{**}	26.26^{***}
C_1^{13}	31.817^{***}												
Panel B													
$\widehat{\alpha}_i$	0.005	0.037	0.093	0.228	0.367	0.489	0.596	0.684	0.764	0.837	0.895	0.927	0.969
$\lvert t_{1,\alpha_i}\rvert$	$\lvert 3.93\rvert^{***}$	$\lvert 5.02\rvert^{***}$	$\lvert 2.08\rvert^{**}$	$\lvert 6.30\rvert^{***}$	$\lvert 12.69\rvert^{***}$	$\lvert 16.98\rvert^{***}$	$\lvert 17.85\rvert^{***}$	$\lvert 16.02\rvert^{***}$	$\lvert 13.51\rvert^{***}$	$\lvert 7.87\rvert^{***}$	$\lvert 1.21\rvert$	$\lvert 7.56\rvert^{***}$	$\lvert 12.54\rvert$
$L^5_{\alpha_i}$	30.30^{***}	73.09^{***}	23.12^{***}	51.01^{***}	193.75^{***}	321.82^{***}	343.19^{***}	284.29^{***}	201.29^{***}	71.28^{***}	7.22	66.05^{***}	198.36^{***}
C_1^{13}	881.66^{***}												
HAR-GARCH model													
Panel C													
$\widehat{\alpha}_i$	0.008	0.049	0.102	0.205	0.311	0.408	0.509	0.611	0.703	0.800	0.899	0.950	0.990
$\lvert t_{1,\alpha_i}\rvert$	$\lvert 1.85\rvert^{*}$	$\lvert 0.23\rvert$	$\lvert 0.74\rvert$	$\lvert 1.45\rvert$	$\lvert 2.95\rvert^{**}$	$\lvert 2.29\rvert^{**}$	$\lvert 2.58\rvert^{**}$	$\lvert 3.45\rvert^{**}$	$\lvert 1.30\rvert$	$\lvert 0.18\rvert$	$\lvert 0.35\rvert$	$\lvert 0.16\rvert$	$\lvert 0.13\rvert$
$L^5_{\alpha_i}$	8.90	4.30	6.09	7.15	12.80^{**}	8.31	9.92^{*}	16.25^{***}	5.12	2.93	8.90	7.08	5.85
C_1^{13}	22.86^{**}												
Panel D													
$\widehat{\alpha}_i$	0.003	0.037	0.099	0.234	0.363	0.480	0.577	0.666	0.749	0.818	0.893	0.930	0.971
$\lvert t_{1,\alpha_i}\rvert$	$\lvert 5.41\rvert^{***}$	$\lvert 4.91\rvert^{***}$	$\lvert 0.26\rvert$	$\lvert 7.44\rvert^{***}$	$\lvert 11.79\rvert^{***}$	$\lvert 14.05\rvert^{***}$	$\lvert 12.88\rvert^{***}$	$\lvert 11.15\rvert^{***}$	$\lvert 8.93\rvert^{***}$	$\lvert 3.62\rvert^{***}$	$\lvert 1.77\rvert^{*}$	$\lvert 6.37\rvert^{***}$	$\lvert 10.70\rvert^{***}$
$L^5_{\alpha_i}$	69.56^{***}	51.28^{***}	8.86	68.17^{***}	152.47^{***}	206.97^{***}	186.68^{***}	147.90^{***}	89.45^{***}	17.17^{***}	10.74^{*}	53.39^{***}	142.98^{***}
C_1^{13}	369.90^{***}												